

BÀI GIẢNG MÔN TIN HỌC ỨNG DỤNG

PHẦN 4

TIN HỌC ỨNG DỤNG TRONG PHÂN TÍCH DỮ LIỆU

Khoa Hệ thống thông tin quản lý
Trường ĐH Ngân hàng TP. HCM



Chương 5

ỨNG DỤNG CÔNG CỤ PHÂN TÍCH DỮ LIỆU

MỤC TIÊU

- *Biết được một số công cụ phân tích dữ liệu thông dụng*
- *Hiểu và áp dụng được thang đo và dữ liệu để giải quyết một số mô hình thực tế*
- *Sử dụng được phần mềm SPSS phục vụ cho phân tích dữ liệu nghiên cứu.*

CÔNG CỤ PHÂN TÍCH DỮ LIỆU

1

GIỚI THIỆU

2

CÁC CÔNG CỤ PHÂN TÍCH DỮ LIỆU

3

ỨNG DỤNG SPSS

1

GIỚI THIỆU

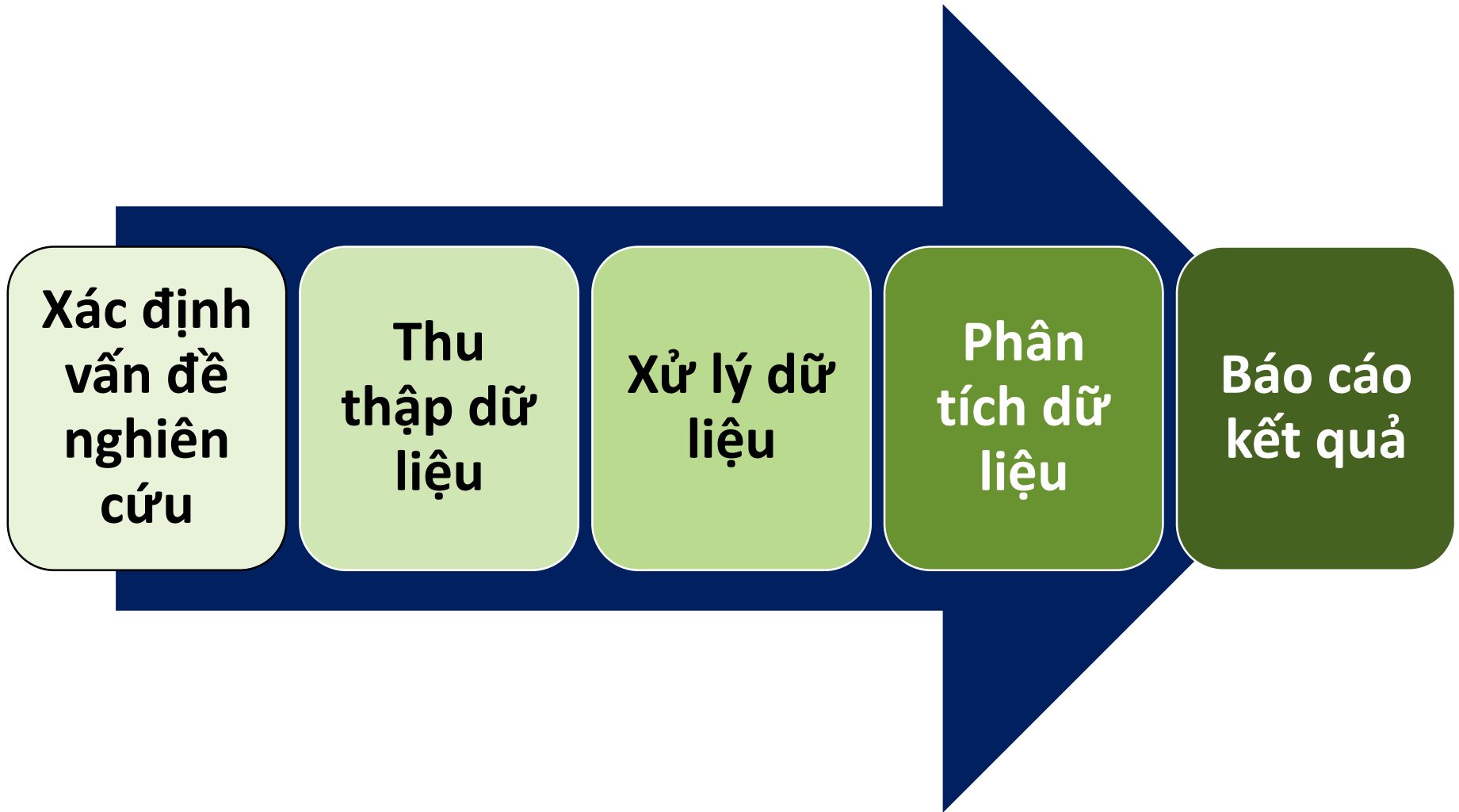
2

CÁC CÔNG CỤ PHÂN TÍCH DỮ LIỆU

3

ỨNG DỤNG SPSS

QUY TRÌNH PHÂN TÍCH DỮ LIỆU



XÁC ĐỊNH VẤN ĐỀ NGHIÊN CỨU

■ Yêu cầu

- Xác định rõ ràng và chính xác mục tiêu, phạm vi, nội dung nghiên cứu
- Xác định (các) biến số cần phân tích, dự báo...

■ Kết quả

- Giúp việc thu thập dữ liệu hiệu quả, chính xác...

THU THẬP DỮ LIỆU

■ Yêu cầu

- Xác định phạm vi tổng thể nghiên cứu, đơn vị điều tra, đơn vị báo cáo
- Xác định mẫu quan sát phục vụ suy diễn thống kê...

■ Kết quả

- Dữ liệu thu thập đáp ứng mục tiêu, phạm vi, nội dung nghiên cứu...

XỬ LÝ DỮ LIỆU

■ Yêu cầu

- Lưu trữ dữ liệu trên máy tính một cách hợp lý, đầy đủ và chính xác
- Loại bỏ sai sót trong thu thập dữ liệu và nhập liệu...

■ Kết quả

- Dữ liệu đầy đủ, chính xác, sẵn sàng cho việc phân tích thống kê...

PHÂN TÍCH DỮ LIỆU

■ Yêu cầu

- Khám phá ý nghĩa thống kê của dữ liệu nghiên cứu
- Xây dựng mối tương quan giữa các biến liên quan đến các vấn đề kinh tế, xã hội trong tương lai...

■ Kết quả

- Kết quả phân tích là chứng cứ thống kê có cơ sở cho việc hiểu biết, gia tăng tri thức và ra quyết định...

BÁO CÁO KẾT QUẢ

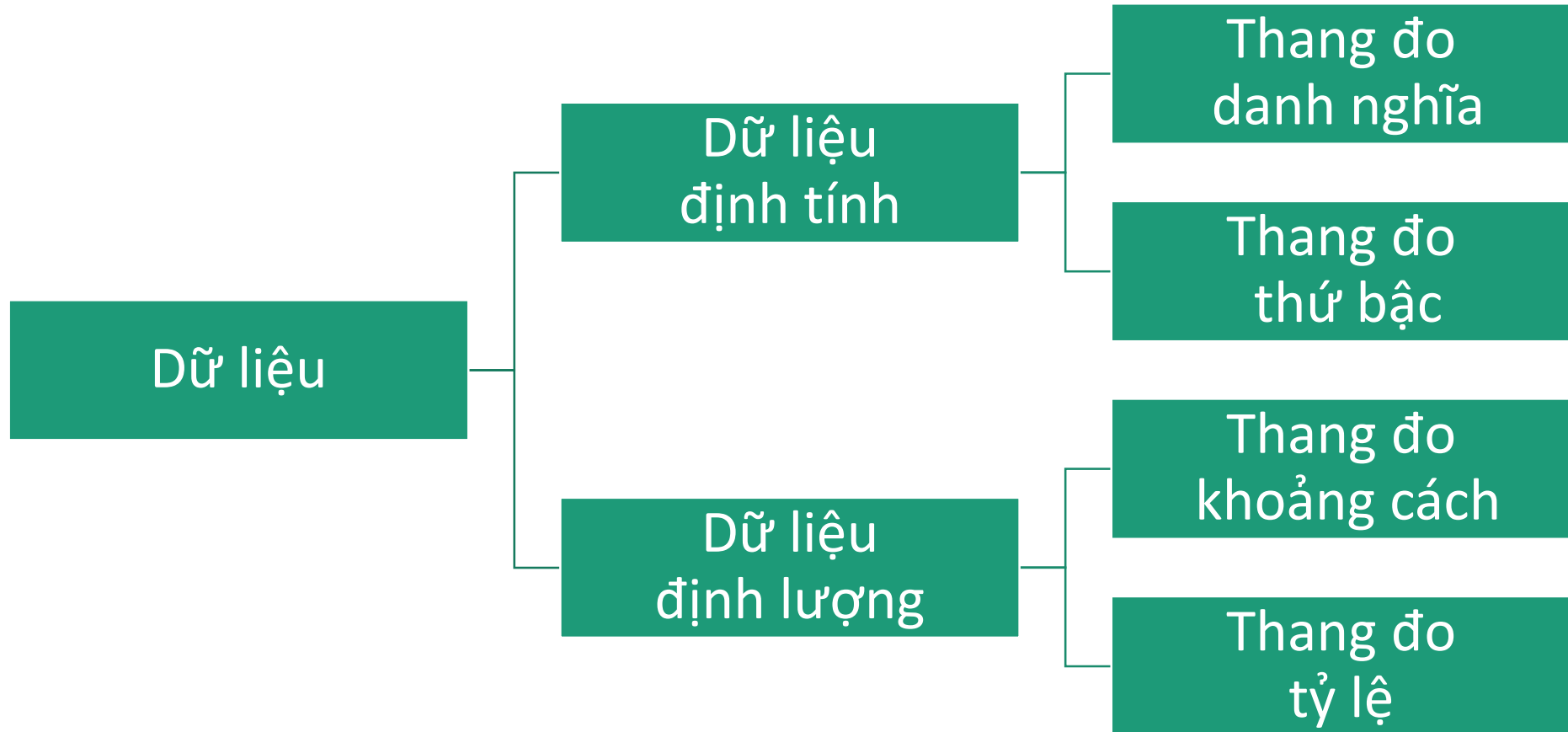
■ Yêu cầu

- Phản ánh kết quả phân tích dữ liệu
- Thể hiện tính không chẵn chẵn của nghiên cứu do phân tích mẫu dữ liệu nghiên cứu hạn chế...

■ Kết quả

- Báo cáo chính xác, đầy đủ, khách quan...

DỮ LIỆU VÀ THANG ĐO



PHÂN LOẠI DỮ LIỆU

▪ Dữ liệu định tính

- Phản ánh tính chất, sự hơn kém
- Thể hiện bằng chuỗi hoặc số
- Không tính được trị trung bình...

▪ Dữ liệu định lượng

- Phản ánh mức độ, mức độ hơn kém
- Thể hiện bằng số chuỗi hoặc số
- Tính được giá trị trung bình...

PHÂN LOẠI DỮ LIỆU

- **Dữ liệu định tính**

- Ví dụ...

- Giới tính
- Tình trạng hôn nhân...

- **Dữ liệu định lượng**

- Ví dụ...

- Thu nhập
- Độ tuổi...



THANG ĐO

- Công cụ mã hóa tình trạng/mức độ của các đơn vị khảo sát theo từng đặc trưng được xem xét
- Thường thực hiện bằng ký số với thứ tự tăng dần từ trên xuống
- Các loại thang đo
 1. Thang đo danh nghĩa (nominal scale)
 2. Thang đo thứ bậc (ordinal scale)
 3. Thang đo khoảng (interval scale)
 4. Thang đo tỷ lệ (ratio scale)...

1. THANG ĐO DANH NGHĨA

- Phân loại đối tượng và đặt tên cho các biểu hiện, ấn định cho chung một ký số tương ứng
- Ý nghĩa
 - Các con số chỉ dùng để phân loại đối tượng
 - Không thể sắp xếp, so sánh...
- Các phép toán thống kê
 - Phép đếm
 - Tính tần suất
 - Xác định giá trị mô hình...

V.D., THANG ĐO DANH NGHĨA

- Câu hỏi khảo sát...
 - Tình trạng nhà ở hiện tại?
 - ☐ Ở nhà thuê
 - ☐ Ở nhà cha mẹ
 - ☐ Ở ký túc xá
 - Các biểu hiện trên có thể được mã hóa...
 - 1 = Ở nhà thuê
 - 2 = Ở nhà cha mẹ
 - 3 = Ở ký túc xá

2. THANG ĐO THỨ BẬC

- Phân loại đối tượng và đặt tên cho các biểu hiện, ấn định cho chung một ký số tương ứng, **được sắp xếp theo một quy ước nào đó**
- Ý nghĩa
 - Các con số được sắp xếp theo thứ bậc/sự hơn kém
 - Không xác định khoảng cách giữa các con số...
- Các phép toán thống kê
 - Số trung vị, số mô hình
 - Khoảng, khoảng tứ trung vị...

V.D., THANG ĐO THỨ BẬC

■ Câu hỏi khảo sát:

- Mức độ hài lòng của khách hàng?

- ☐ Hài lòng
- ☐ Bình thường
- ☐ Không hài lòng

- Cách biểu hiện trên có thể được quy ước

3 = Hài lòng

2 = Bình thường

1 = Không hài lòng

3. THANG ĐO KHOẢNG

- Phân loại đối tượng và đặt tên cho các biểu hiện, ấn định cho chung một ký số tương ứng, sắp xếp với một khoảng cách nhất định giữa các giá trị
- Ý nghĩa
 - Các con số được sắp xếp theo thứ bậc/sự hơn kém
 - Xác định khoảng cách giữa các con số
- Các phép toán thống kê
 - Số trung vị, số mô hình, khoảng, khoảng tứ trung vị
 - Khoảng biến thiên, số trung bình, độ lệch chuẩn
 - Có thể thực hiện tính (+, -); không hỗ trợ chia (/)...

V.D., THANG ĐO KHOẢNG

■ Câu hỏi khảo sát:

- Tầm quan trọng của các yếu tố sau đây đối với chất lượng đào tạo đại học?

Các thành phần	Không quan trọng			Rất quan trọng	
	1	2	3	4	5
Chương trình đào tạo					
Cơ sở vật chất					
Phương pháp giảng dạy					
Đội ngũ giáo viên					
Công tác hỗ trợ sinh viên					

4. THANG ĐO TỶ LỆ

- Phân loại đối tượng và đặt tên cho các biểu hiện, ấn định cho chung một ký số tương ứng, được sắp xếp với một khoảng cách nhất định giữa các giá trị
- Ý nghĩa
 - Các con số được sắp xếp theo thứ bậc/sự hơn kém
 - Xác định khoảng cách giữa các con số
- Các phép toán thống kê
 - Số trung vị, số mode, khoảng, khoảng tứ trung vị
 - Khoảng biến thiên, số trung bình, độ lệch chuẩn
 - Hỗ trợ phép tính (+, -, /)...

V.D., THANG ĐO TỶ LỆ

- Câu hỏi khảo sát:

- Độ tuổi?

- Người 40 tuổi gấp 2 lần tuổi so với người 20 tuổi, nhưng chỉ bằng $\frac{2}{3}$ lần tuổi so với người 60 tuổi...

- Thu nhập bình quân?

- Người thu nhập 20 triệu/tháng gấp đôi so với người thu nhập 10 triệu/tháng...

MÃ HÓA SỐ LIỆU

- Mục đích

- Chuyển đổi thông tin đã thu thập thành dạng thích hợp cho việc phân tích trên máy tính

- Nguyên tắc

- Thông tin từng đối tượng thể hiện trong một hàng
- Mỗi cột ứng với một câu hỏi trong công cụ thu thập
- Mỗi câu trả lời được mã hóa trên một ô của bảng tính
- Câu trả lời được mã hóa bằng số, hạn chế bằng chữ...

V.D: MÃ HÓA SỐ LIỆU

1. Anh/chị bao nhiêu tuổi (tính tròn năm)?

2. Anh/chị là nam hay nữ (khoanh vào lựa chọn phù hợp)?

Nam

Nữ

3. Anh/chị đã học đến lớp mấy (Khoanh vào lựa chọn phù hợp)?

Không đi học

Tiểu học

Dưới THCS

Trên THCS

Học nghề

Đại học

Trên đại học

Trẻ nhỏ

V.D: MÃ HÓA SỐ LIỆU

Mô tả	Giá trị - ý nghĩa
Tuổi	1 - 01-30 2 - 31-60 3 - 61-99...
Giới tính	1 - Nam 2 - Nữ 3 - Khác...
Trình độ học vấn	1 - Không biết chữ 2 - Tiểu học 3 - Trung học cơ sở 4 - Trung học phổ thông 5 - Trung cấp nghề 6 - Đại học 7 - Sau đại học 9 - Không trả lời...

NHẬP SỐ LIỆU

■ Mục đích

- Cung cấp bộ dữ liệu phục vụ cho việc phân tích

■ Nguyên tắc

- Kiểm tra sự lặp lại hoặc các giá trị không có ý nghĩa
- Hạn chế đến mức thấp nhất các lỗi nhập liệu
- Chọn chiến lược nhập liệu phù hợp và làm sạch dữ liệu

■ Phương pháp

- Nhập liệu, kiểm tra và xử lý một số lỗi trên Excel
- Chuyển dữ liệu từ Excel sang SPSS để phân tích...

CÁC CHIẾN LƯỢC NHẬP LIỆU

- Nhập toàn bộ số liệu 2 lần bởi 2 người độc lập
- Nhập toàn bộ số liệu 2 lần bởi cùng 1 người
- Nhập số liệu 1 lần, và nhập 20% ngẫu nhiên lần 2
- Nhập toàn bộ số liệu 1 lần d



1

GIỚI THIỆU

2

CÁC CÔNG CỤ PHÂN TÍCH DỮ LIỆU

3

ỨNG DỤNG SPSS

CÁC PHẦN MỀM THÔNG DỤNG

- MS Excel
- SPSS
- EViews
- Stata
- R...



MS EXCEL

- MS Excel sử dụng rộng rãi trong kế toán, phân tích thống kê, phân tích kinh doanh...
- Đặc điểm...
 - Thực hiện nhiều phép tính phức tạp
 - Thư viện hàm phong phú, chuyên sâu về thống kê
 - Tích hợp nhiều công cụ phân tích, xử lý số liệu
 - Chức năng quản trị cơ sở dữ liệu, biểu đồ...

CÁC CÔNG CỤ EXCEL

- Goal Seek
- Data Table
- Solver
- Scenario Manager
- Analysis ToolPak
- Currency Tool
- VBA...



SPSS

- **SPSS** (Statistical Package for the Social Sciences)
 - Phần mềm thống kê được sử dụng rộng rãi trong các nghiên cứu điều tra xã hội học và kinh tế lượng
- Đặc điểm...
 - Phát triển từ 1968 tại Đại học Stanford, thương mại từ 1975, IBM mua lại 2009...
 - Hoạt động trên Windows, Mac OS, Linux, Unix...
 - Phiên bản mới nhất (02/2018): **SPSS 25**...

ƯU VÀ NHƯỢC ĐIỂM CỦA SPSS

■ Ưu điểm

- Giao diện thân thiện, dễ sử dụng
- Thực hiện thao tác tính toán đơn giản
- Hỗ trợ xử lý và phân tích dữ liệu
- Có thể mạnh về phân tích nhân tố, phương sai, hồi quy...
- Khả năng lập bảng biểu dữ liệu, báo cáo đa dạng, linh hoạt

■ Nhược điểm

- Không có khả năng lập trình
- Không cho phép xử lý cùng lúc nhiều tập tin dữ liệu
- Thiếu một số chức năng phân tích thống kê chuyên sâu...

EViews

- EViews (Econometric Views)

- Phần mềm kinh tế lượng

- Đặc điểm...

- Sử dụng các dữ liệu có sẵn - dữ liệu thứ cấp
 - Dùng cho dữ liệu chuỗi, dữ liệu chéo, dữ liệu mảng...
 - Thao tác linh hoạt, quản lý dễ dàng, kết quả nhanh
 - Phiên bản mới nhất (02/2018): EViews 10...

CÁC ỨNG DỤNG EVIEWS

- Thống kê mô tả dữ liệu
- Phân tích tác động của các yếu tố kinh tế
- Dự báo cho tương lai yếu tố cần nghiên cứu
- Thực hiện xếp hạng tín dụng trong ngân hàng
- Tính giá trị rủi ro cổ phiếu dựa trên tỉ suất sinh lời
- Phân tích tương quan giữa các yếu tố kinh tế...

STATA

- STATA (Data Analysis and Statistical Software)
 - Phần mềm sử dụng trong kinh tế lượng và thống kê
- Đặc điểm...
 - Kết hợp tính dễ sử dụng với sức mạnh thống kê
 - Thực hiện thao tác bằng lệnh, soạn thảo chương trình
 - Không cho phép xử lý cùng lúc nhiều tập tin dữ liệu
 - Mạnh về phân tích hồi qui
 - Phiên bản mới nhất (02/2018): Stata 16...

CÁC ỨNG DỤNG STATA

- Quản lý dữ liệu
- Phân tích thống kê
- Vẽ đồ thị
- Mô phỏng
- Hiệu chỉnh chức năng...



R



1

GIỚI THIỆU

2

CÁC CÔNG CỤ PHÂN TÍCH DỮ LIỆU

3

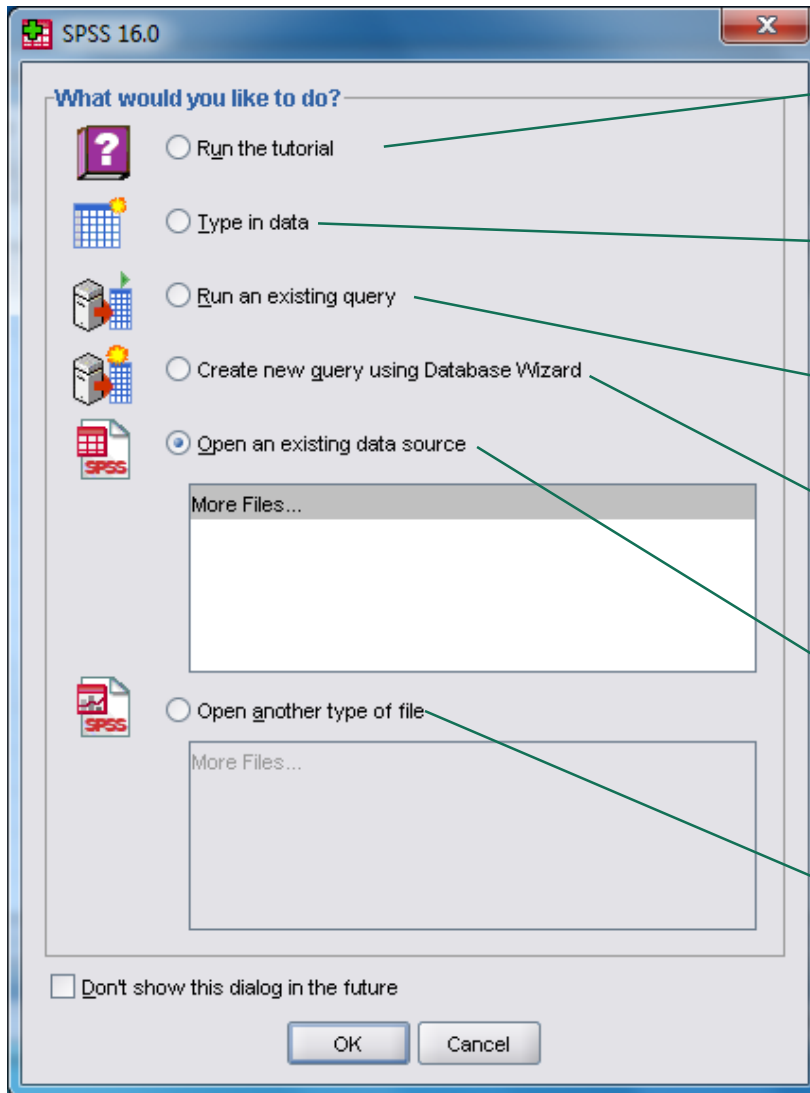
ỨNG DỤNG SPSS

TỔNG QUAN VỀ SPSS

- Khởi động SPSS
- Giao diện SPSS
- Tập tin SPSS
- Làm việc với Data
- Làm việc với Viewer
- Tiếng Việt trong SPSS
- Thay đổi mặc định của chương trình...



KHỞI ĐỘNG SPSS



Mở hướng dẫn sử dụng

Mở Data Editor

Thực hiện Script có sẵn

Tạo Script mới

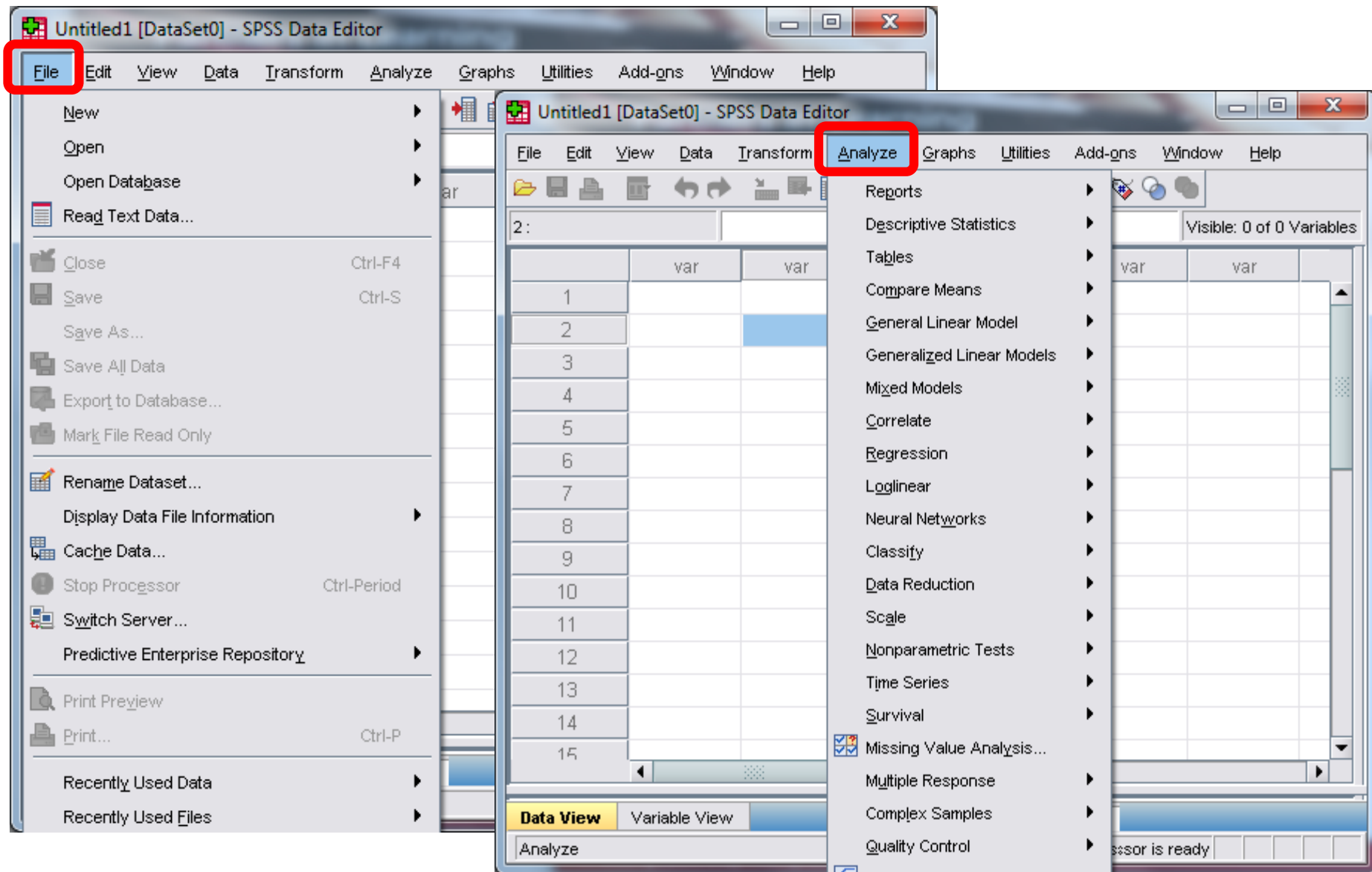
Mở file dữ liệu có sẵn

Mở file không phải SPSS

GIAO DIỆN SPSS

- **Data Editor** dạng bảng tính để định nghĩa, nhập, hiệu chỉnh và thể hiện dữ liệu...
- **Viewer** thể hiện kết quả phân tích dữ liệu
- **Multidimension pivot table** kết quả dạng bảng trụ
- **High revolution graphics** kết quả phân tích dạng đồ thị
- **Database access** truy xuất dữ liệu CSDL bên ngoài SPSS
- **Data transformation** biến đổi dữ liệu gốc cho phù hợp với mục đích nghiên cứu...

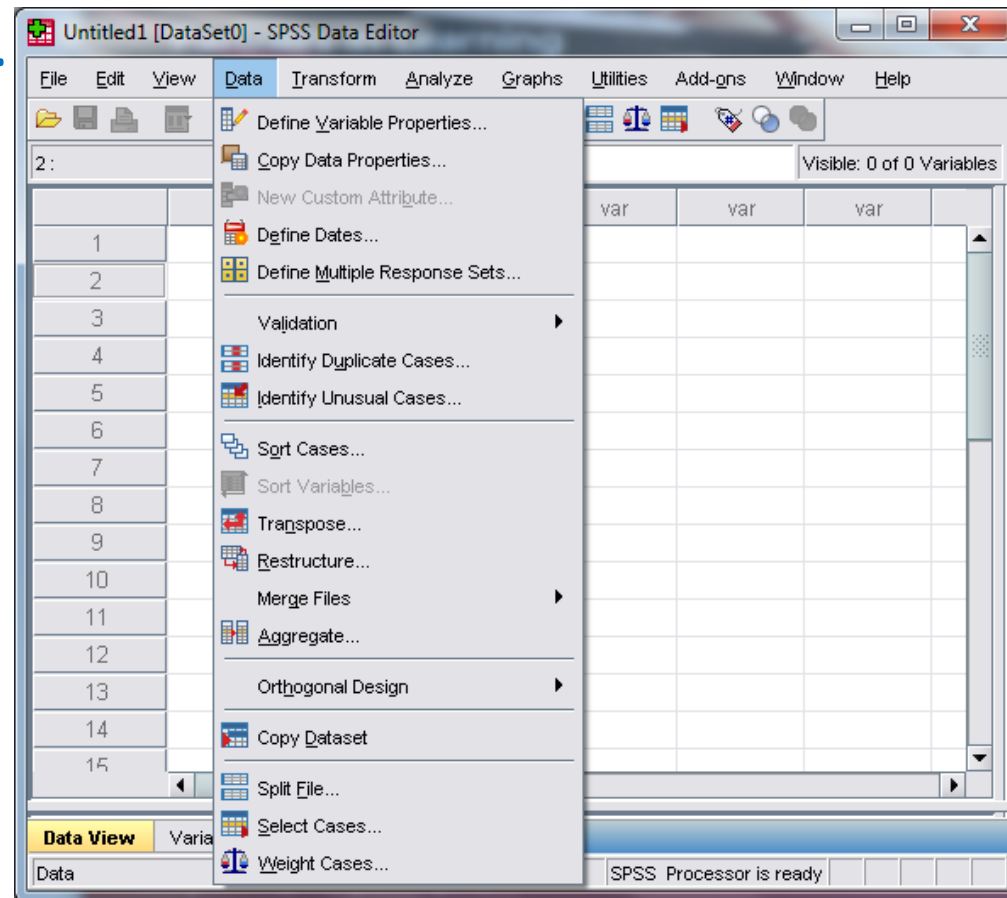
V.D., GIAO DIỆN SPSS



GIAO DIỆN SPSS (THỰC ĐƠN)

■ Data

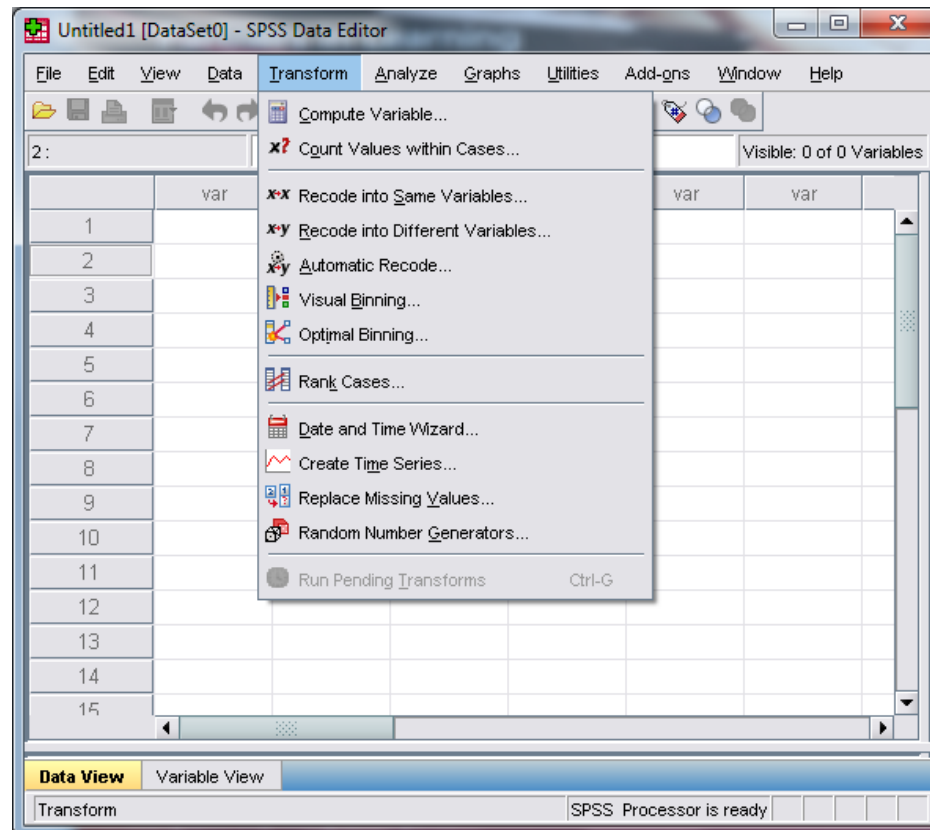
- Định nghĩa đặc trưng của biến; kiểm tra tính hợp lệ của dữ liệu; sắp xếp, biến đổi biến/quan sát; phân tách/kết hợp các tập tin dữ liệu...



GIAO DIỆN SPSS (THỰC ĐƠN)

■ Transform

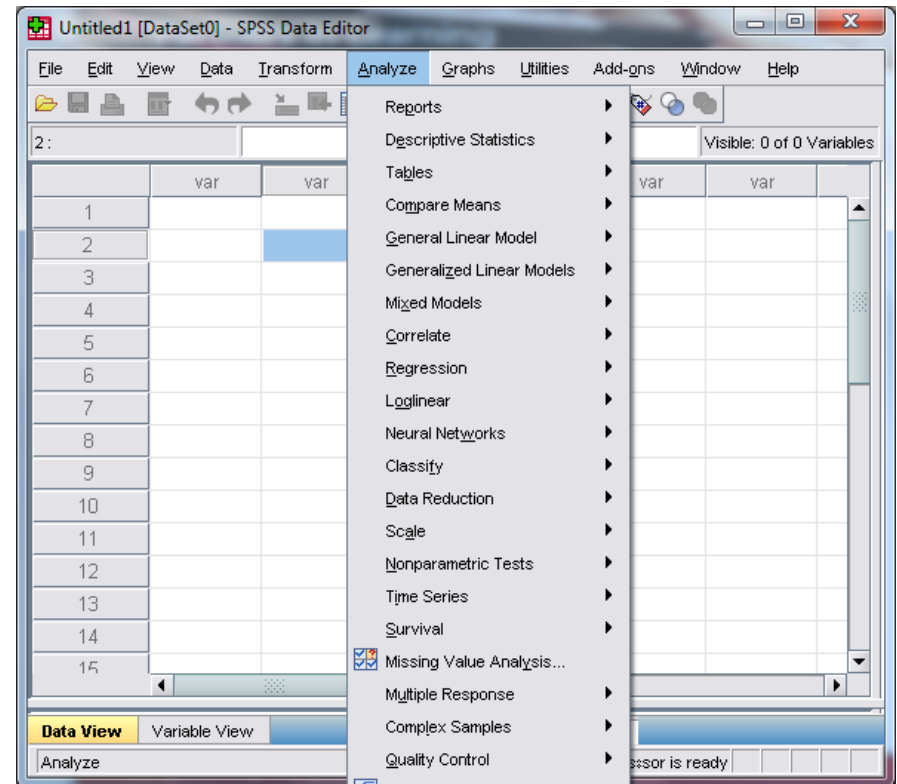
- Tạo biến mới từ biến có sẵn; thiết lập các chuỗi dữ liệu thời gian; xử lý các trường hợp quan sát thiếu dữ liệu...



GIAO DIỆN SPSS (THỰC ĐƠN)

■ Analyze

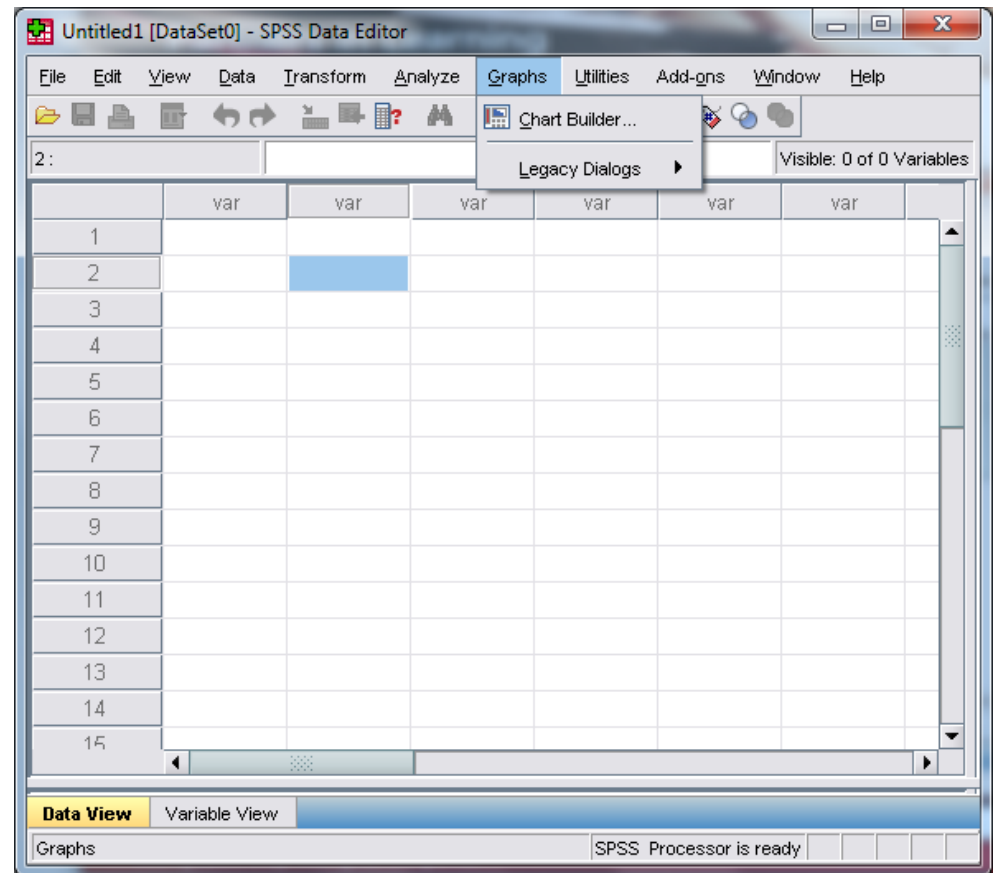
- Các công cụ phân tích dữ liệu (phân tích thống kê mô tả, kiểm định trung bình tổng thể, phân tích tương quan, phân tích hồi quy, kiểm định tham số/phi tham số...)



GIAO DIỆN SPSS (THỰC ĐƠN)

■ Graphs

- Xây dựng các loại biểu đồ



LÀM VIỆC VỚI TẬP TIN SPSS

- Tạo mới tập tin SPSS
- Mở tập tin SPSS có sẵn
- Lưu tập tin SPSS...

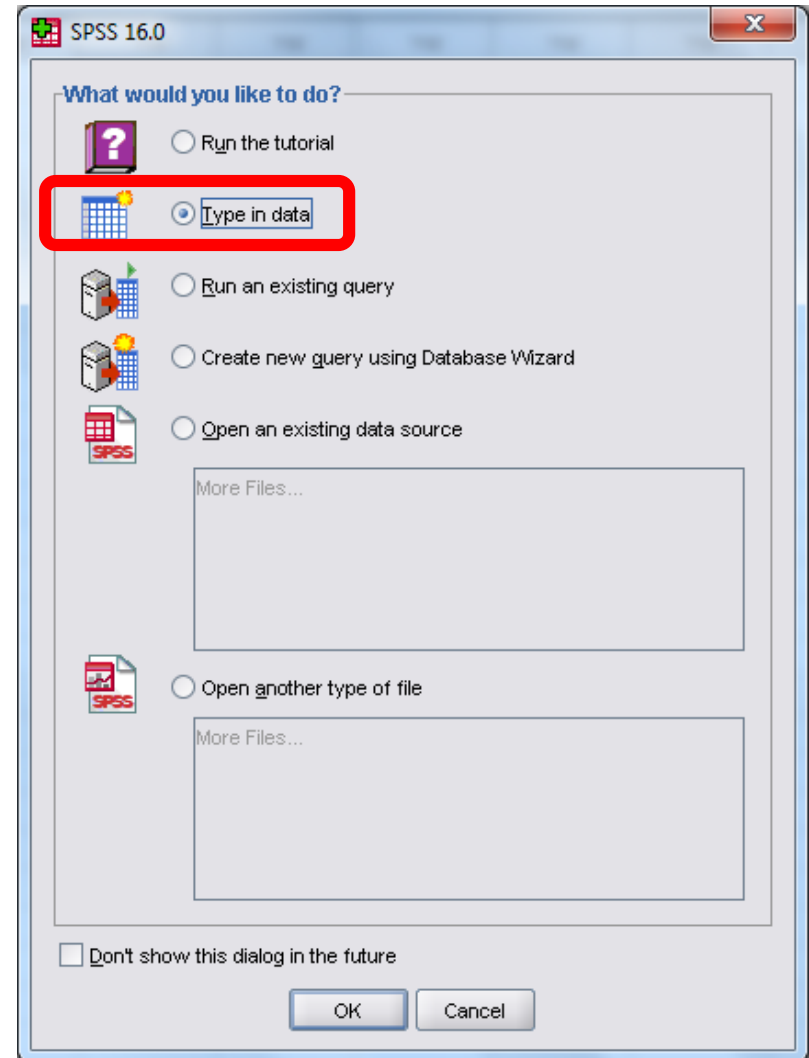
Spss1.sav... Spss2.spv

CÁC LOẠI TẬP TIN DỮ LIỆU SPSS

- Các định dạng tập tin SPSS
 - Dữ liệu: *.sav hoặc *.sys
 - Cú pháp: *.sps
 - Kết quả: *.spv
 - Script: *.wwd hoặc *.sbs.
- SPSS còn đọc tập tin từ các định dạng khác...
 - *.xls (Excel)
 - *.txt (Text)
 - *.dta (Stata)
 - *.wf1 (EViews)...

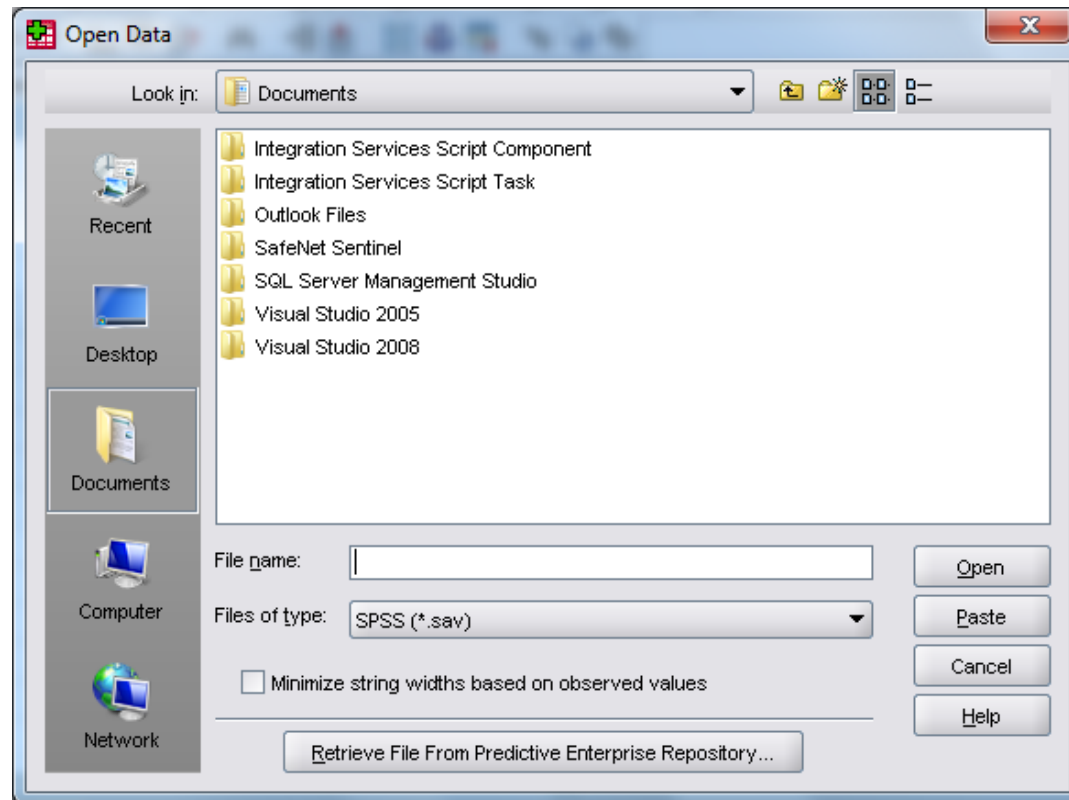
TẠO TẬP TIN MỚI

- Khởi động SPSS
→ **Type in Data** →
OK
- Từ cửa sổ hiện có
SPSS
→ **File** → **New** →
Data

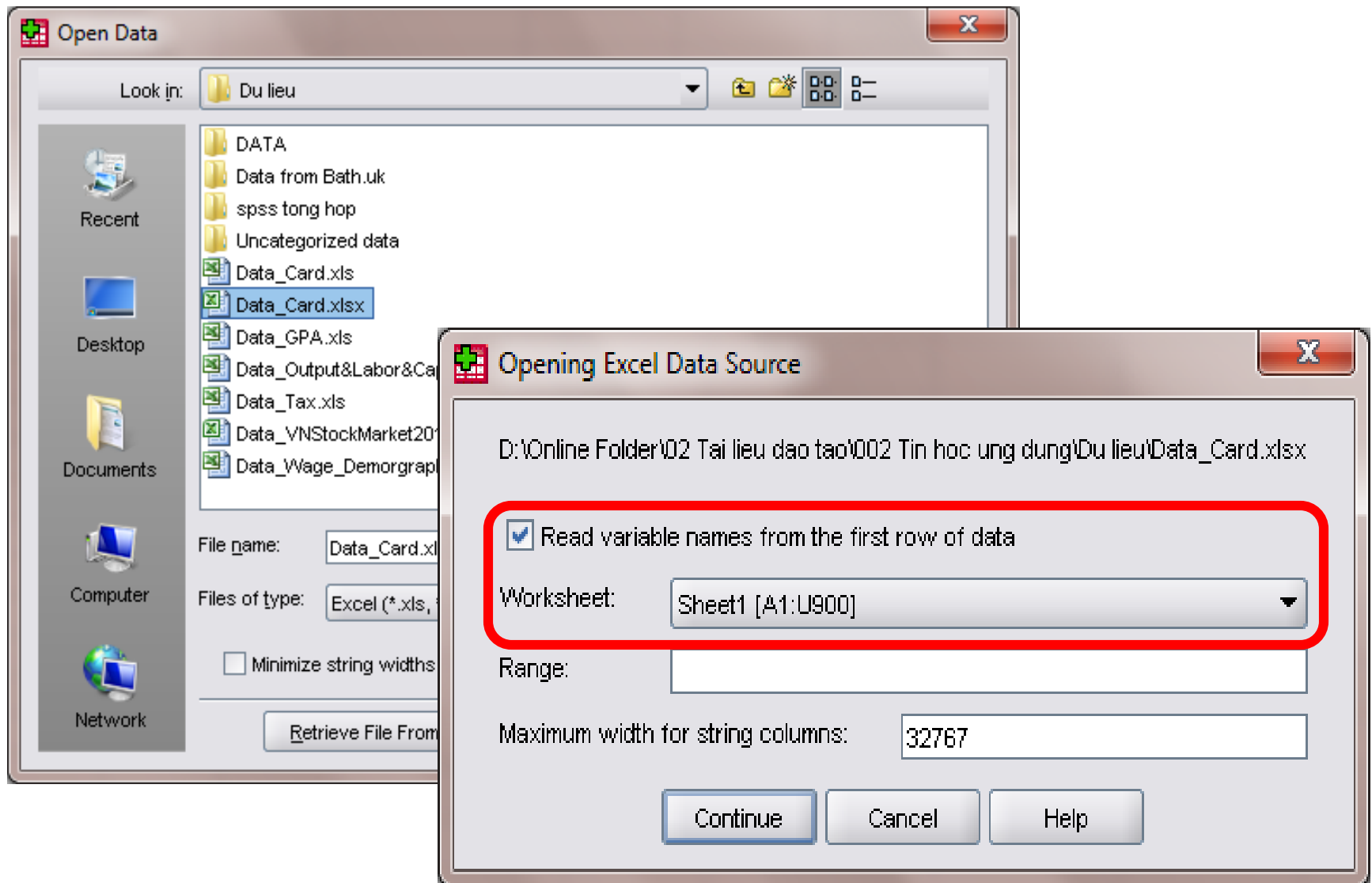


MỞ TẬP TIN CÓ SẴN

- Nhấn hai lần vào biểu tượng của tập tin SPSS
- Khởi động SPSS → **Open an existing** → **OK**
- Từ cửa sổ hiện có SPSS → **File** → **Open** → **Data...**

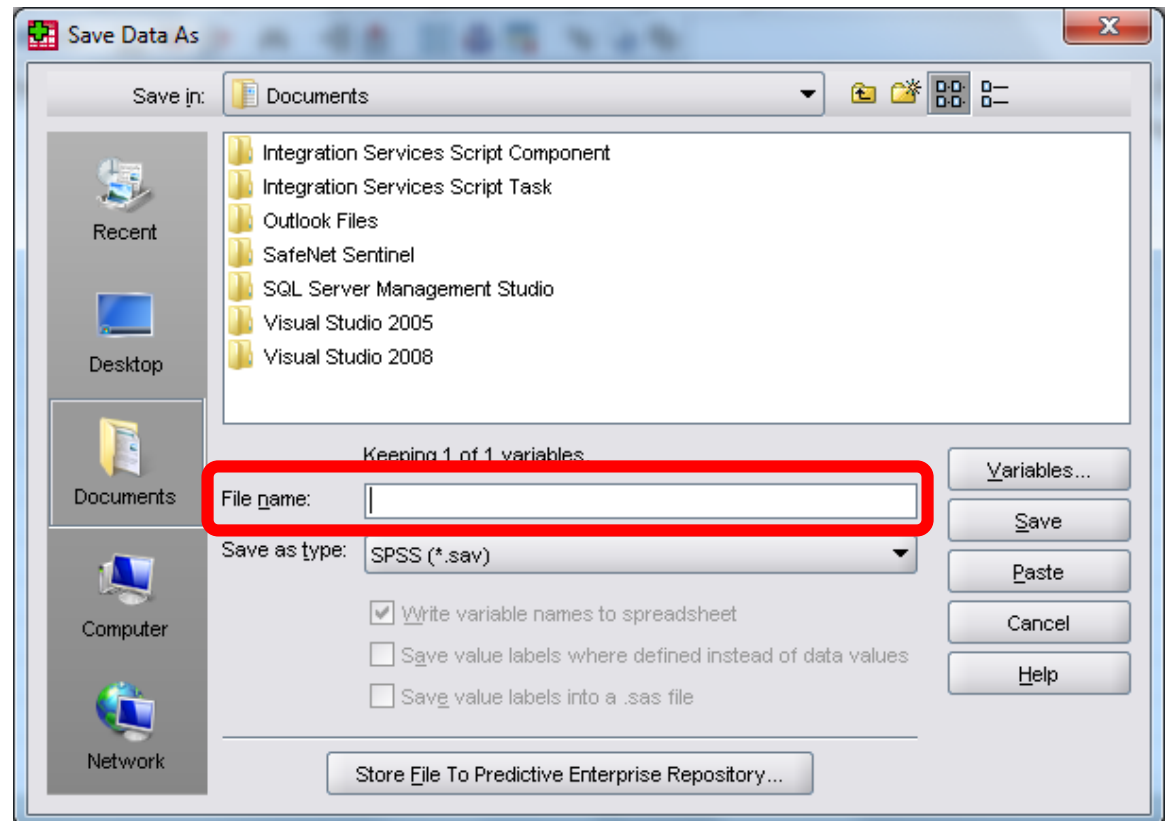


V.D., MỞ TẬP TIN TỪ EXCEL



LƯU TẬP TIN

- Tập tin đang mở → **Ctrl + S...**
- Tập tin mới → **File → Save → Save File As** → tên tập tin...



LÀM VIỆC VỚI DATA EDITOR

- Data Editor thể hiện như một bảng tính:
 - Data View thể hiện trị số dữ liệu thực hoặc các nhãn trị số được xác định
 - Variable View thể hiện thông tin định nghĩa biến
 - Nhãn biến, nhãn trị số biến, loại dữ liệu, thang đo, và các trị số khuyết thiếu...

DATA VIEW

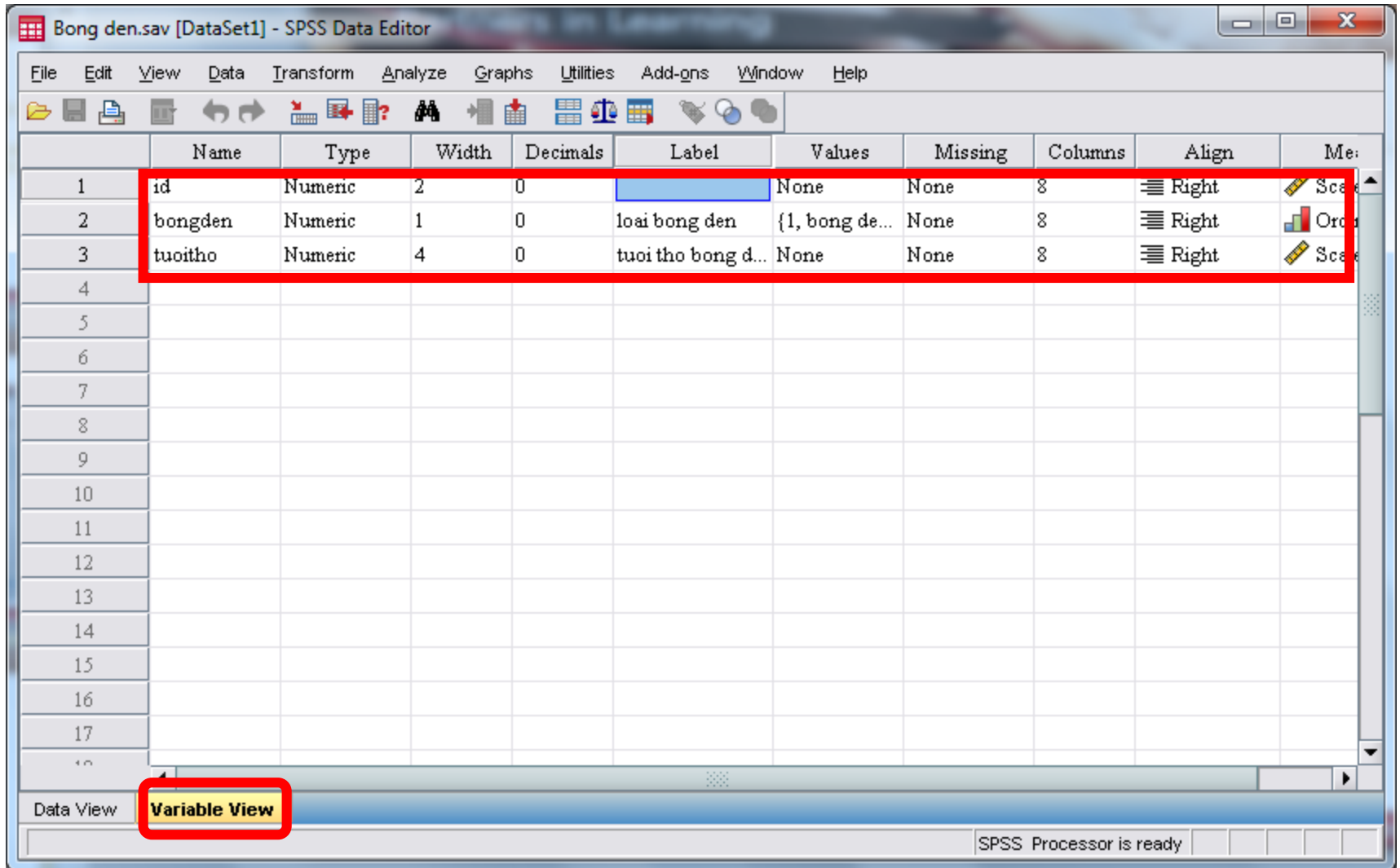
SPSS Data Editor window titled "Bong den.sav [DataSet1]". The interface shows a menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, Help) and a toolbar. The main area displays a data grid with 16 rows and 10 columns. The first four columns are labeled "id", "bongden", "tuaitho", and "var". The first column "id" is highlighted in blue. A red box highlights the first 16 rows of data. The status bar at the bottom shows "Data View" selected, "Variable View" as an alternative, and "SPSS Processor is ready".

	id	bongden	tuaitho	var	var	var	var	var	var
1	1	1	2400						
2	2	1	3800						
3	3	1	2300						
4	4	1	2600						
5	5	1	2400						
6	6	1	2800						
7	7	1	2100						
8	8	2	3900						
9	9	2	3200						
10	10	2	2900						
11	11	2	3400						
12	12	2	2700						
13	13	3	3000						
14	14	3	2100						
15	15	3	1800						
16	16	2	2200						

HIỆU CHỈNH DỮ LIỆU - DATA VIEW

- Thay đổi trị số của dữ liệu
- Cắt, sao chép, dán các trị số
- Thêm/xóa các đối tượng
- Thêm/xóa các biến
- Thay đổi trật tự các biến...

VARIABLE VIEW



CÁC QUY TẮC KHAI BÁO BIẾN

- Tên biến
- Kiểu dữ liệu
- Nhãn trị số của biến...

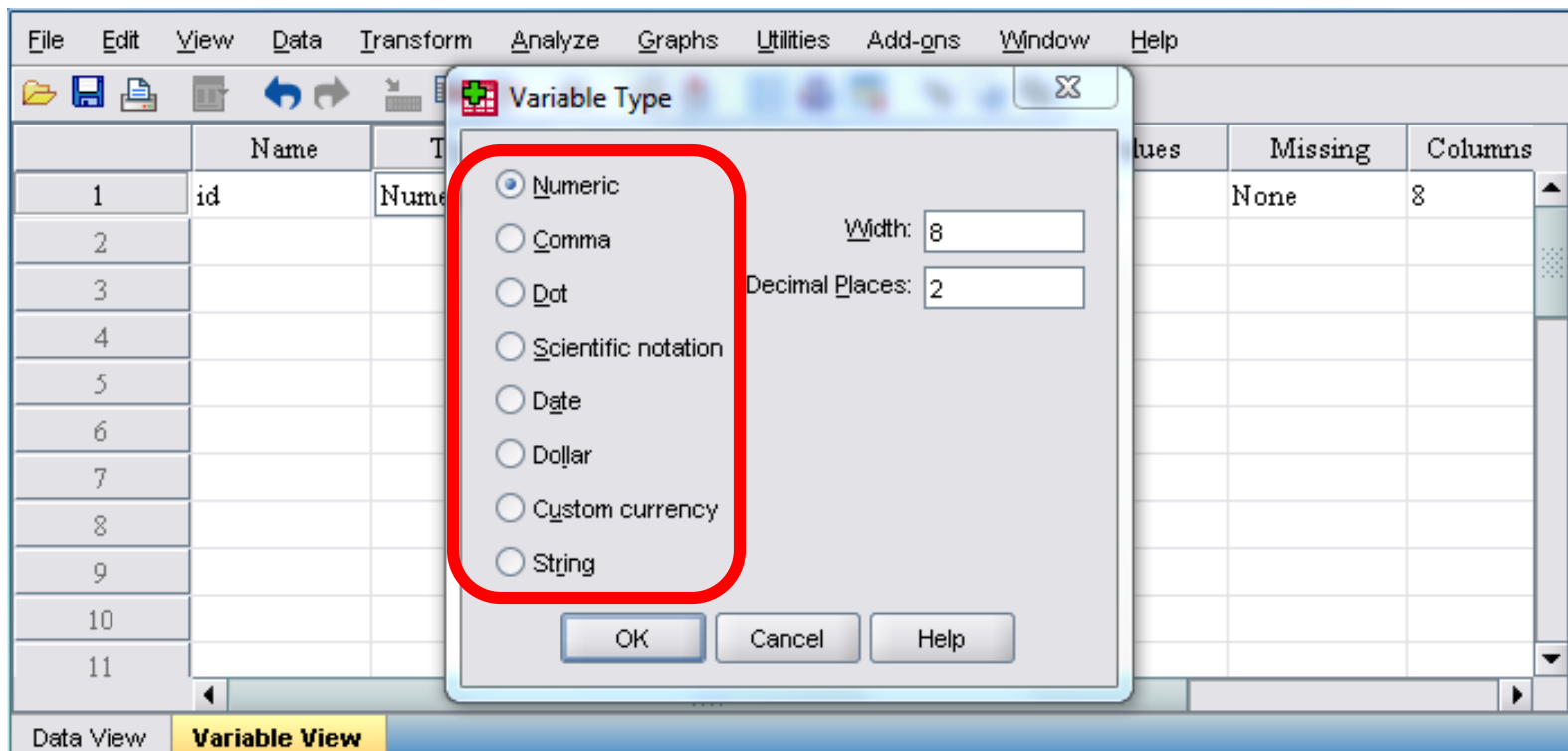
Name... Label

TÊN BIẾN

- Tên bắt đầu bằng một ký chữ, các ký tự còn lại có thể là chữ, số, hoặc ký tự đặc biệt (@, #, _, \$), và không kết thúc bởi dấu chấm (.)
- Biến không trùng lặp, không phân biệt chữ hoa/thường
- Độ dài biến không quá 8 ký tự...

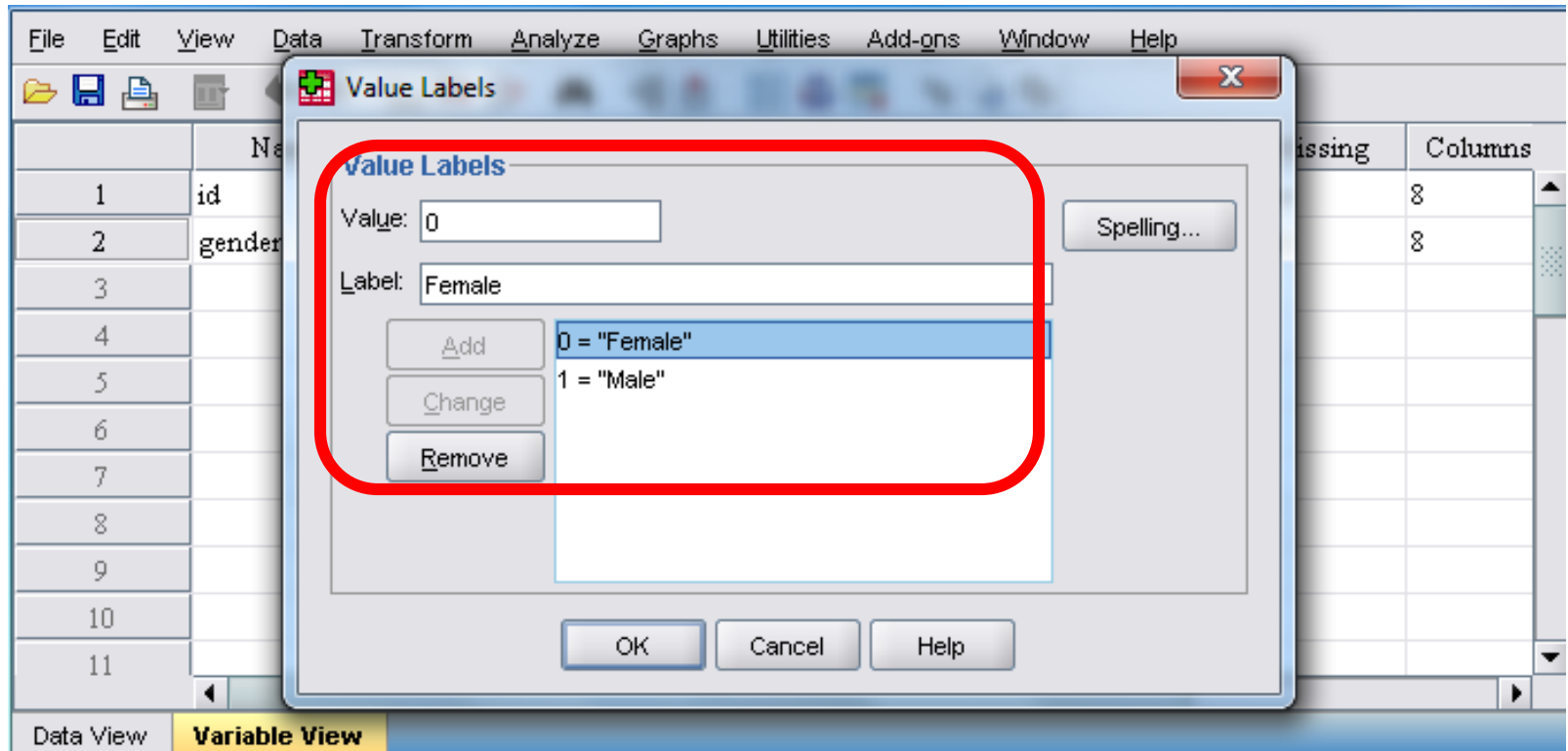
KIỂU DỮ LIỆU

- Xác định loại dữ liệu với từng biến
- Mặc định là kiểu số (chiều dài 8, 2 số thập phân)...



NHÃN TRỊ SỐ CỦA BIẾN

- Chỉ định nhãn mô tả đối với từng trị số của biến
- Nhãn chỉ số của biến có chiều dài tối đa 60...



LÀM VIỆC VỚI VIEWER

■ Mục đích

- Thể hiện kết quả ứng với thao tác/thủ tục trên dữ liệu

■ Các tính năng của Viewer

- Hỗ trợ xem kết quả thao tác trên dữ liệu
- Điều khiển cách thức hiển thị kết quả
- Lưu kết quả tài liệu theo tổ chức, định dạng thích hợp...

V.D: SPSS VIEWER

The screenshot shows the SPSS Viewer window titled '*Output1 [Document1] - SPSS Viewer'. The window has a menu bar (File, Edit, View, Data, Transform, Insert, Format, Analyze, Graphs, Utilities, Add-ons, Window, Help) and a toolbar. On the left is a tree view of the output structure:

- Output
 - Log
 - Frequencies
 - Title
 - Notes
 - Active Dataset
 - Statistics
 - Frequency Table
 - Title
 - loai bong den
 - tuoi tho bong den (gio)

The main pane displays the following syntax script:

```
GET  
  FILE='D:\02 Tai lieu dao tao\002 Tin hoc ung dung\Phan tich du lieu  
nghien cuu voi SPSS - 2008 Data Full\Bong den.sav'.  
DATASET NAME DataSet0 WINDOW=FRONT.  
FREQUENCIES VARIABLES=bongden tuoi tho  
  /ORDER=ANALYSIS.
```

Below the script, the output is shown:

[DataSet1] D:\02 Tai lieu dao tao\002 Tin hoc ung dung\Phan tich du lieu

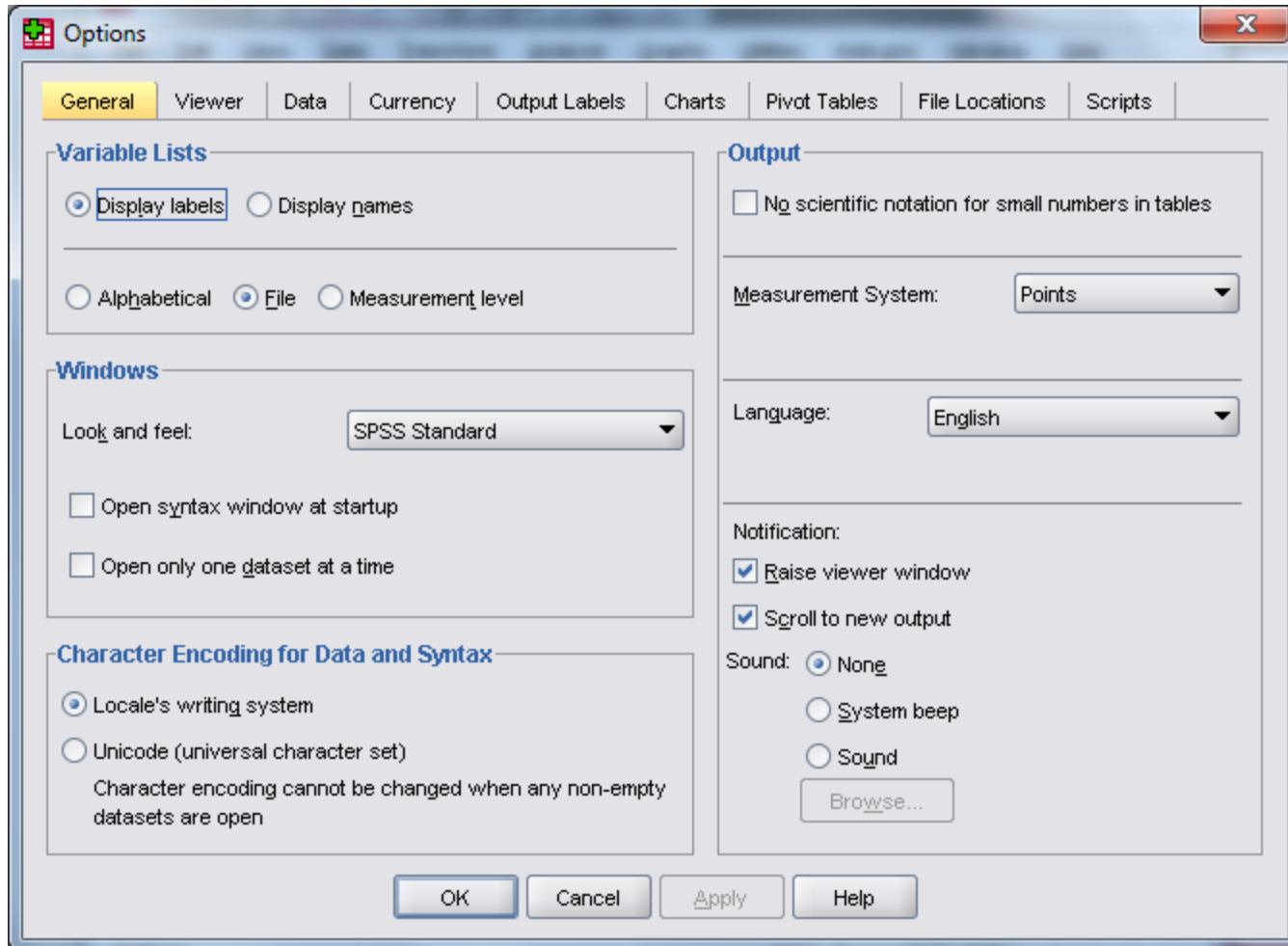
A red arrow points from the 'Statistics' node in the tree view to the following table:

		loai bong den	tuoi tho bong den (gio)
N	Valid	17	17
	Missing	0	0

The status bar at the bottom indicates 'SPSS Processor is ready'.

THAY ĐỔI MẶC ĐỊNH CHƯƠNG TRÌNH

- **Edit → Options → Thay đổi thông số → Apply → OK...**



XỬ LÝ DỮ LIỆU TRÊN BIẾN

1. Mã hóa biến với **Recode**
2. Chuyển định dạng của biến với **Count value**
3. Tính toán giá trị của biến với **Compute**
4. Xử lý câu hỏi có nhiều lựa chọn trả lời...

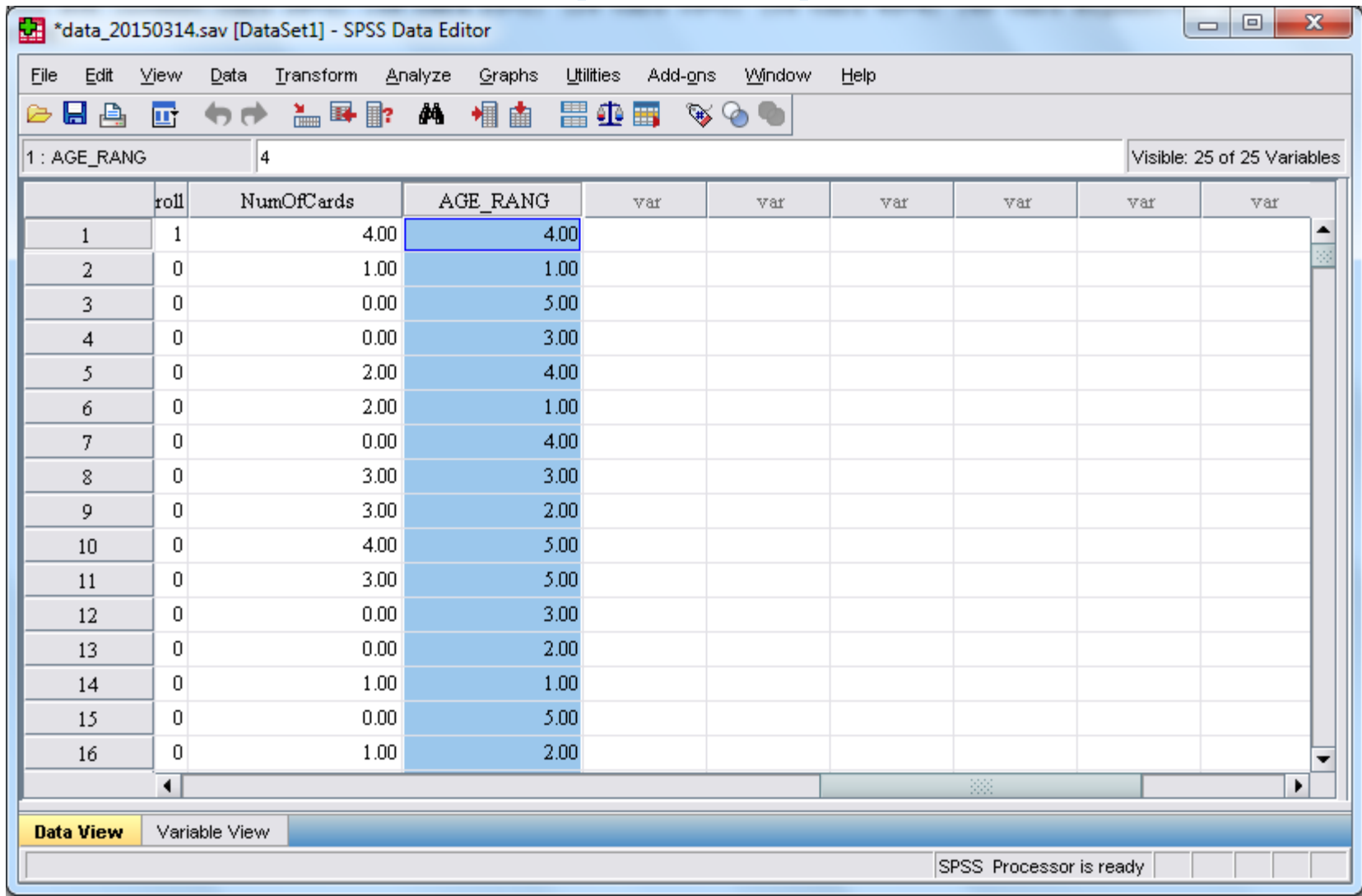
1. MÃ HÓA BIẾN (RECODE)

- Mục đích
 - Biến đổi trị số dữ liệu bằng cách mã hóa lại
- Điều kiện áp dụng
 - Giảm số biểu hiện của 1 biến định tính xuống chỉ còn 2/3 loại biểu hiện cơ bản
 - Chuyển biến định lượng thành biến định tính.

MÃ HÓA BIẾN (TT)

- Quy trình thực hiện
 - **Transform → Recode Into Different Variables...**
 - Chọn biến muốn mã hóa lại
 - Đặt tên cho biến mới → **Name → Change...**
 - Giá trị cho biến cũ và mới → **Old and New Values → → Add → Continue → OK...**

V.D: MÃ HÓA BIẾN VỚI RECODE (SPSS)



The screenshot shows the SPSS Data Editor window for a file named *data_20150314.sav. The 'Data View' tab is active, displaying a dataset with 16 rows and 10 columns. The first column is labeled 'roll' and contains values from 1 to 16. The second column is labeled 'NumOfCards' and contains values from 0.00 to 4.00. The third column is labeled 'AGE_RANG' and contains values from 1.00 to 5.00. The remaining columns are labeled 'var' and are empty. The status bar at the bottom indicates 'SPSS Processor is ready'.

	roll	NumOfCards	AGE_RANG	var	var	var	var	var	var
1	1	4.00	4.00						
2	0	1.00	1.00						
3	0	0.00	5.00						
4	0	0.00	3.00						
5	0	2.00	4.00						
6	0	2.00	1.00						
7	0	0.00	4.00						
8	0	3.00	3.00						
9	0	3.00	2.00						
10	0	4.00	5.00						
11	0	3.00	5.00						
12	0	0.00	3.00						
13	0	0.00	2.00						
14	0	1.00	1.00						
15	0	0.00	5.00						
16	0	1.00	2.00						

V.D., THIẾT LẬP GIÁ TRỊ BIẾN MỚI

*data_20150314.sav [DataSet1] - SPSS Data Editor

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

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	
18	AGR	Numeric	3	0		None	None	3	Left	
19	SACOM	Numeric	5	0		None	None	5	Left	
20	OTHER	Numeric	5	0		None	None	5	Left	
21	PAYMENT	Numeric	2	0		None	None	7	Left	
22	RANGE_OF...	Numeric	8	0		{1, 1-18}...	None	10	Right	
23	VCBPayroll	Numeric	8	0	VCB Payroll	None	None	6	Right	
24	NumOfCards	Numeric	8	2		None	None	12	Right	
25	AGE_RANG	Numeric	8	0		{1, 1-18}...	None	10	Right	
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										

Data View Variable View

SPSS Processor is ready

2. CHUYỂN ĐỊNH DẠNG BIẾN (COUNT VALUE)

- Mục đích
 - Chuyển biến dạng Category nhiều trị số thành biến dạng Dichotomy có 2 trị số...

- Ví dụ

- *Payroll biến phân loại, cho biết tên ngân hàng quản lý tài khoản lương của người được hỏi*
- *ABC_Payroll biến phân loại, cho biết người được hỏi có nhận lương qua ngân hàng ABC hay không...*

CHUYỂN ĐỊNH DẠNG BIẾN (TT)

- Quy trình thực hiện
 - **Transform** → *Count Value within Cases...*
 - Khai báo tên và nhãn của biến muốn phân loại
 - Cung cấp **Category** có liên quan tới **Dichotomy**
 - **Define Value** → Cung cấp trị số mã hóa của **Category** có liên quan tới **Dichotomy**
 - Gán tên các giá trị của biến **Dichotomy**...

CHUYỂN ĐỊNH DẠNG BIẾN VỚI COUNT VALUE

*data_20150314.sav [DataSet1] - SPSS Data Editor

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

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
18	AGR	Numeric	3	0		None	None	3	Left
19	SACOM	Numeric	5	0		None	None	5	Left
20	OTHER	Numeric	5	0		None	None	5	Left
21	PAYMENT	Numeric	2	0		None	None	7	Left
22	RANGE_OF...	Numeric	8	0		{1, 1-18}...	None	10	Right
23	VCBPayroll	Numeric	8	0	VCB Payroll	None	None	6	Right
24	NumOfCards	Numeric	8	2		None	None	12	Right
25	AGE_RANG	Numeric	8	0		{1, 1-18}...	None	10	Right
26	ACBPayroll	Numeric	8	0	ACBPayroll	None	None	12	Right
27									
28									
29									
30									
31									
32									
33									
34									
35									

Data View **Variable View**

SPSS Processor is ready

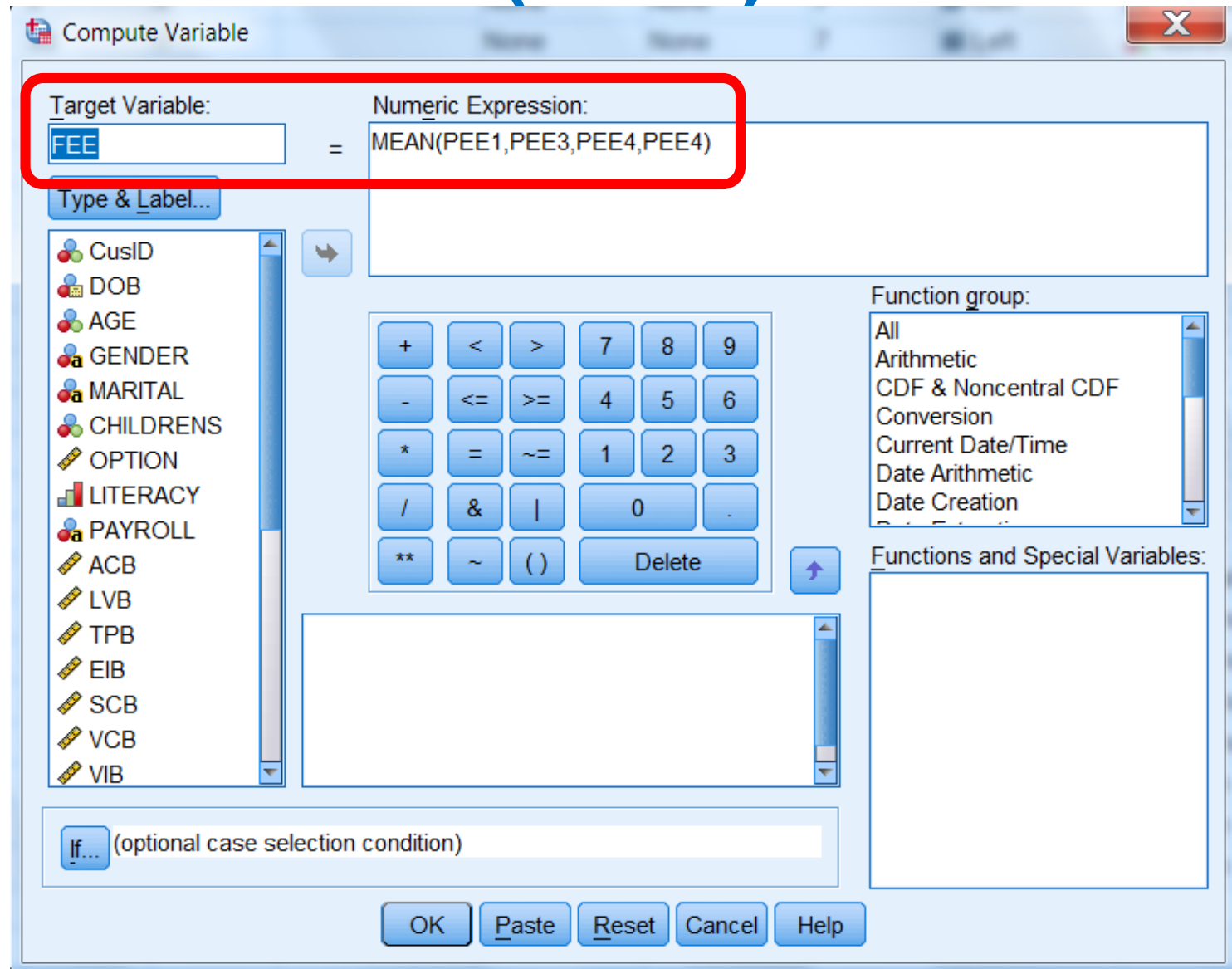
3. TÍNH TOÁN GIÁ TRỊ (COMPUTE)

- Mục đích
 - Rút ngắn thời gian nhập liệu và gán giá trị cho biến
- Điều kiện áp dụng
 - Tạo biến mới không điều kiện
 - Tạo biến mới từ các cấu trúc của các biến đã
- Ví dụ
 - Tính giá trị trung bình của các biến *FE1, FE2, FE3, FE4...*

TÍNH TOÁN GIÁ TRỊ (TT)

- Quy trình thực hiện
 - **Transform** → *Compute...*
 - Chọn các biến muốn tính toán
 - **Target Variable** → **Type & Label** đặt tên biến mới
 - **Numeric Expression** nhập giá trị gán cho biến mới
 - **If** → **Include if case satisfies condition** để thiết đặt điều kiện biến...

TÍNH TOÁN GIÁ TRỊ VỚI COMPUTE (SPSS)



4. CÂU HỎI NHIỀU LỰA CHỌN TRẢ LỜI

- Mục đích
 - Mã hóa câu hỏi có nhiều lựa chọn trả lời
- Điều kiện áp dụng
 - Đây là trường hợp phức tạp và cần thực hiện nhiều thao tác với các câu hỏi có nhiều lựa chọn trả lời
- Ví dụ
 - Thống kê mô tả theo loại hình ngân hàng (1, 2, 3, 4).

CÂU HỎI NHIỀU LỰA CHỌN TRẢ LỜI (TT)

- Quy trình thực hiện (b1)
 - Nhập liệu cho đáp ứng có nhiều chọn (v.d., 1 2 3 4...)
 - **Analyze → Multiple Response → Define Variable...**
 - Chọn giá các trị **Set Define → Variable in Set...**
 - Chọn **Variables are...** và **Range... Through...** → **Name** và **Label** → **Add** giá trị vào **Multiple Response Sets...**

CÂU HỎI NHIỀU LỰA CHỌN TRẢ LỜI (SPSS)

Define Multiple Response Sets

Set Definition

Variables in Set:

- C1.1
- C1.2
- C1.3
- C1.4
- C1.5
- C1.6
- C1.7

Variables Are Coded As:

☐ Dichotomies Counted value:

☒ Categories Range: through

Name: CAU1 ← Nhập tên biến mới vào đây

Label: CAU1 ← Nhập Label hoặc để trống

Multiple Response Sets:

Add **Change** **Remove**

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.

CÂU HỎI NHIỀU LỰA CHỌN TRẢ LỜI (TT)

- Quy trình thực hiện (b2)
 - **Analyze → Multiple Response → Frequencies**
 - **Chọn Table(s) for... → OK**

\$CAU1 Frequencies				
		Responses		Percent of Cases
		N	Percent	
CAU1 ^a	CAU 1.1	67	25.1%	33.5%
	CAU 1.2	158	59.2%	79.0%
	CAU 1.3	6	2.2%	3.0%
	CAU 1.4	36	13.5%	18.0%
Total		267	100.0%	133.5%
a. Group				

IBM

SPSS®

THỐNG KÊ MÔ TẢ

1. Tần số - **Frequencies**
2. Thông số thống kê mô tả - **Descriptives**
3. Thống kê mô tả - **Explorer**

1. TẦN SỐ (FREQUENCIES)

- Mục đích

- Thống kê số lượng đối tượng theo từng biểu hiện của thuộc tính

- Quy trình thực hiện

- **Analyze → Descriptive Statistics → Frequencies**
- Chọn biến muốn thống kê trong cửa sổ Frequencies

V.D: TÍNH TẦN SỐ (RANGE-AGE)

The screenshot shows the SPSS Data Editor window with the 'Analyze' menu open and the 'Frequencies' dialog box displayed. The 'Frequencies' dialog box has 'RANGE_OF_AGE' selected in the 'Variable(s):' list. The 'OK' button is highlighted. The background data table is partially visible.

	CusID	DOB	VCB	PAYROLL	ACB	LVB	TPB	EIB	SCB
1	101	02/21/1974			0	0	0	1	1
2	102	04/				1	0	0	0
3	103	10/				0	0	0	0
4	104	12/				0	0	0	0
5	105	12/				0	0	0	1
6	106	08/				1	0	0	0
7	107	02/				0	0	0	0
8	108	04/				0	0	0	1
9	109	11/				0	0	0	1
10	110	07/				0	0	1	1
11	111	05/				0	1	1	0
12	112	03/				0	0	0	0
13	113	12/				0	0	0	0
14	114	08/				0	0	0	0
15	115	02/20/1955				0	0	0	0

V.D., BẢNG TẦN SỐ (SPV)

% hợp lệ của từng biểu hiện

Tần suất của từng biểu hiện

Tần số của từng biểu hiện

Các biểu hiện của biến

% tích lũy của từng biểu hiện

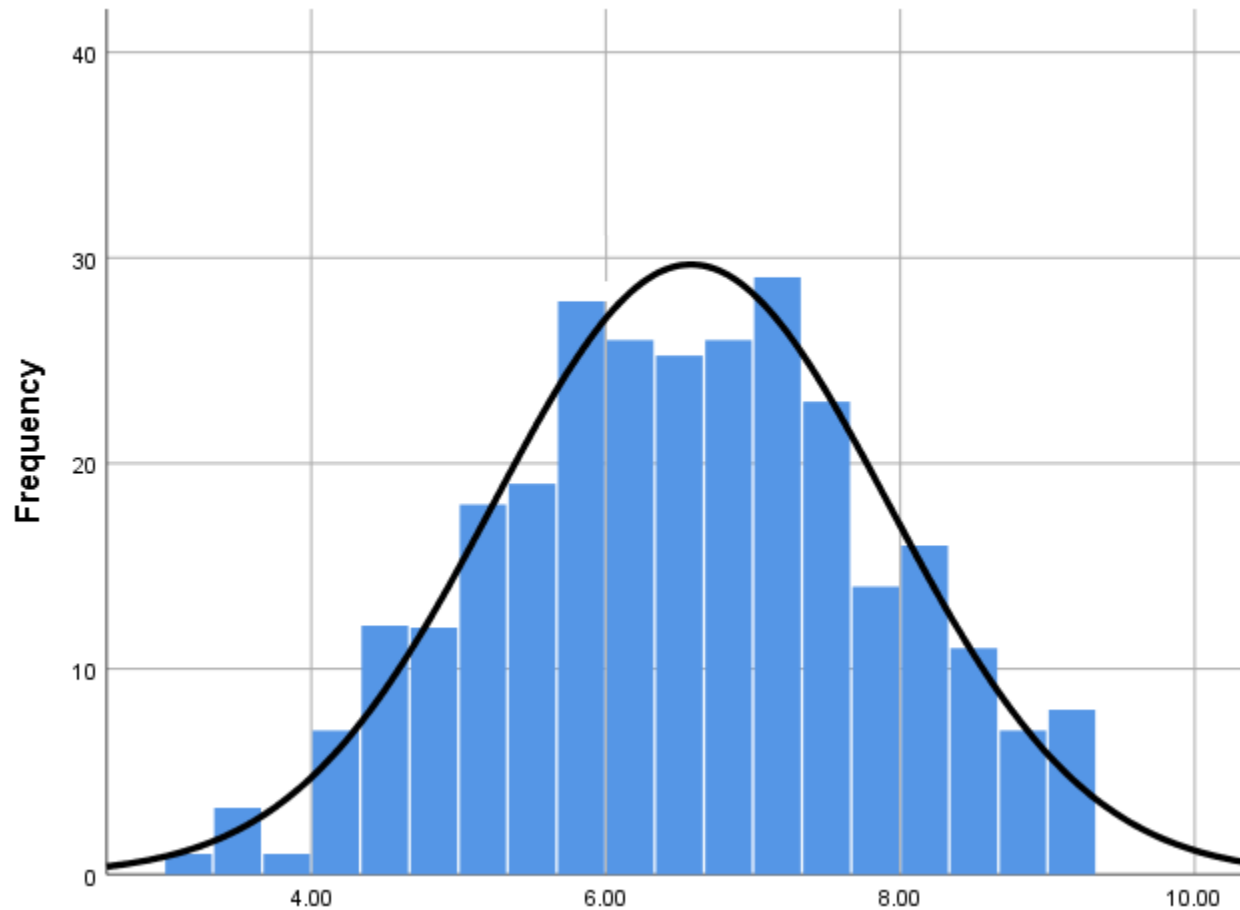
Số quan sát hợp lệ

Số quan sát bị thiếu dữ liệu

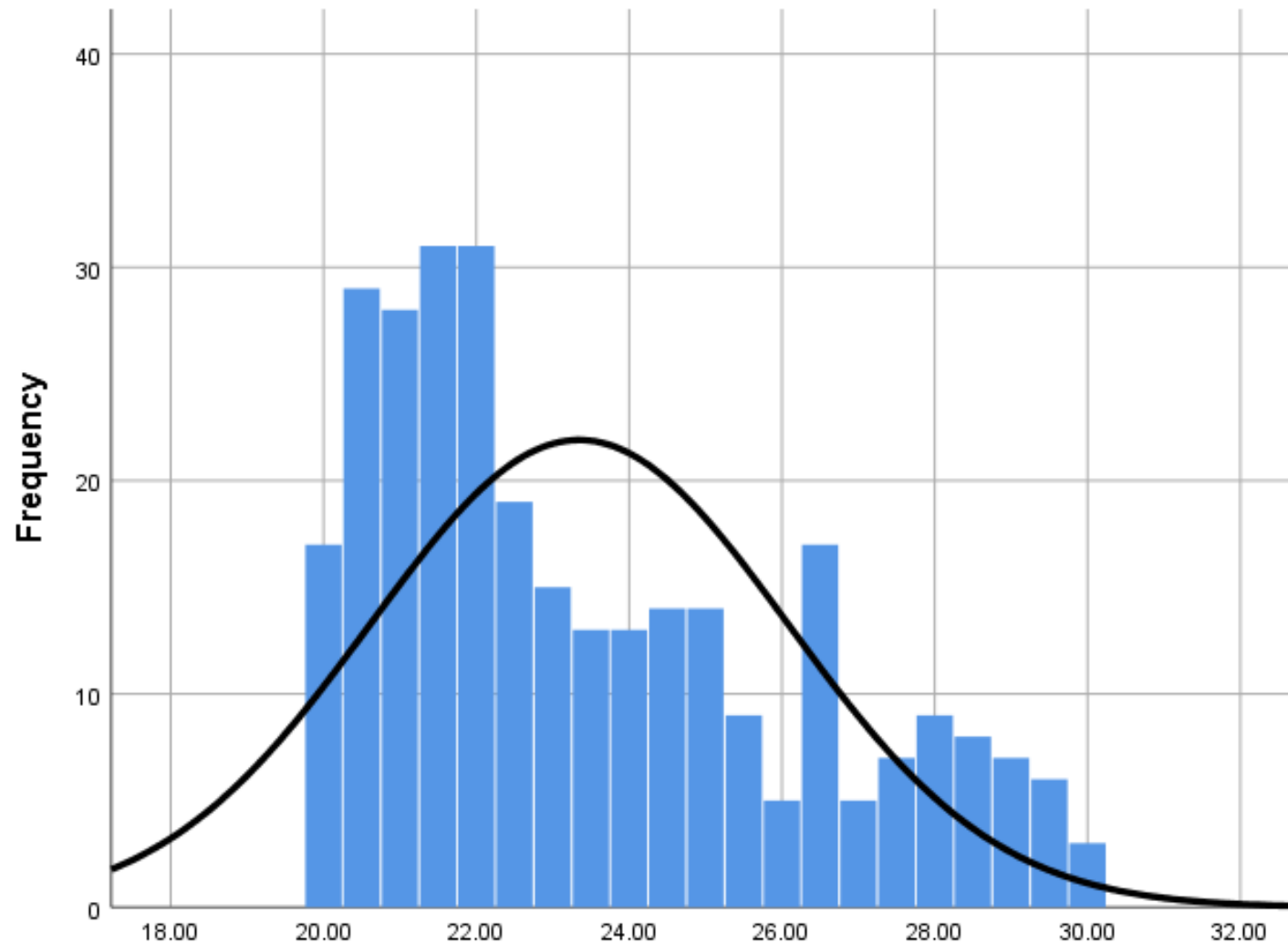
Statistics		
RANGE_OF_AGE		
N	Valid	899
	Missing	0

RANGE_OF_AGE				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
1-18	72	8.0	8.0	8.0
18-25	110	12.2	12.2	20.2
25-36	208	23.1	23.1	43.4
36-45	178	19.8	19.8	63.2
45-60	331	36.8	36.8	100.0
Total	899	100.0	100.0	

PHÂN PHỐI CHUẨN (SPV)

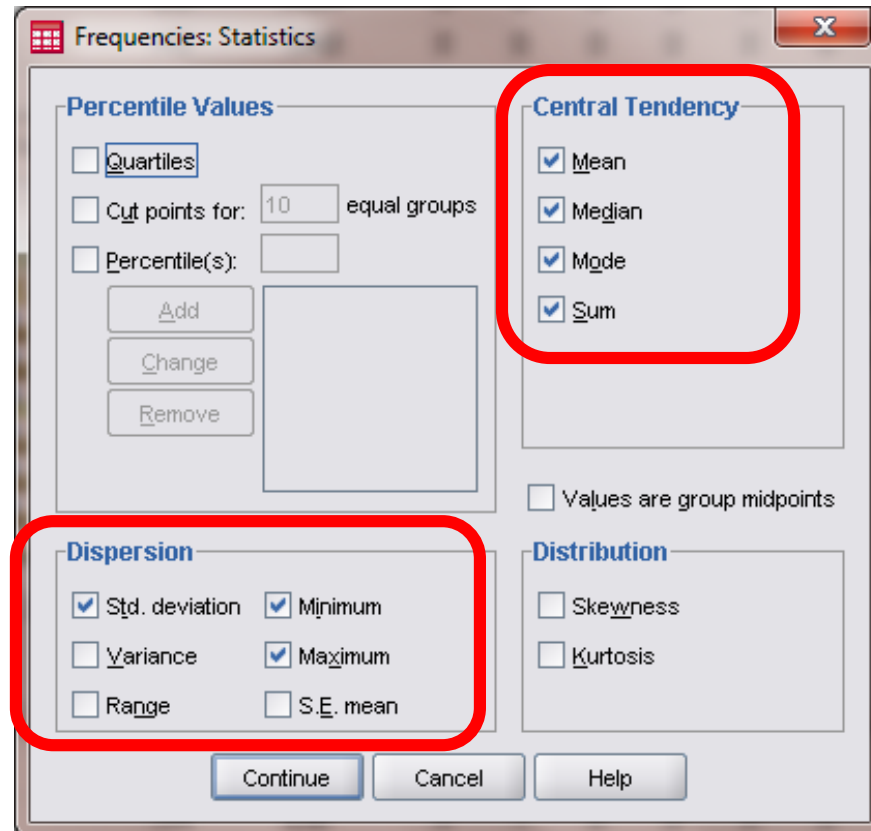


PHÂN PHỐI KHÔNG CHUẨN (SPV)



2. CÁC THÔNG SỐ THỐNG KÊ MÔ TẢ

- Max
- Min
- Median
- Mean
- Var
- Stdev



THỐNG KÊ MÔ TẢ - FREQUENCIES

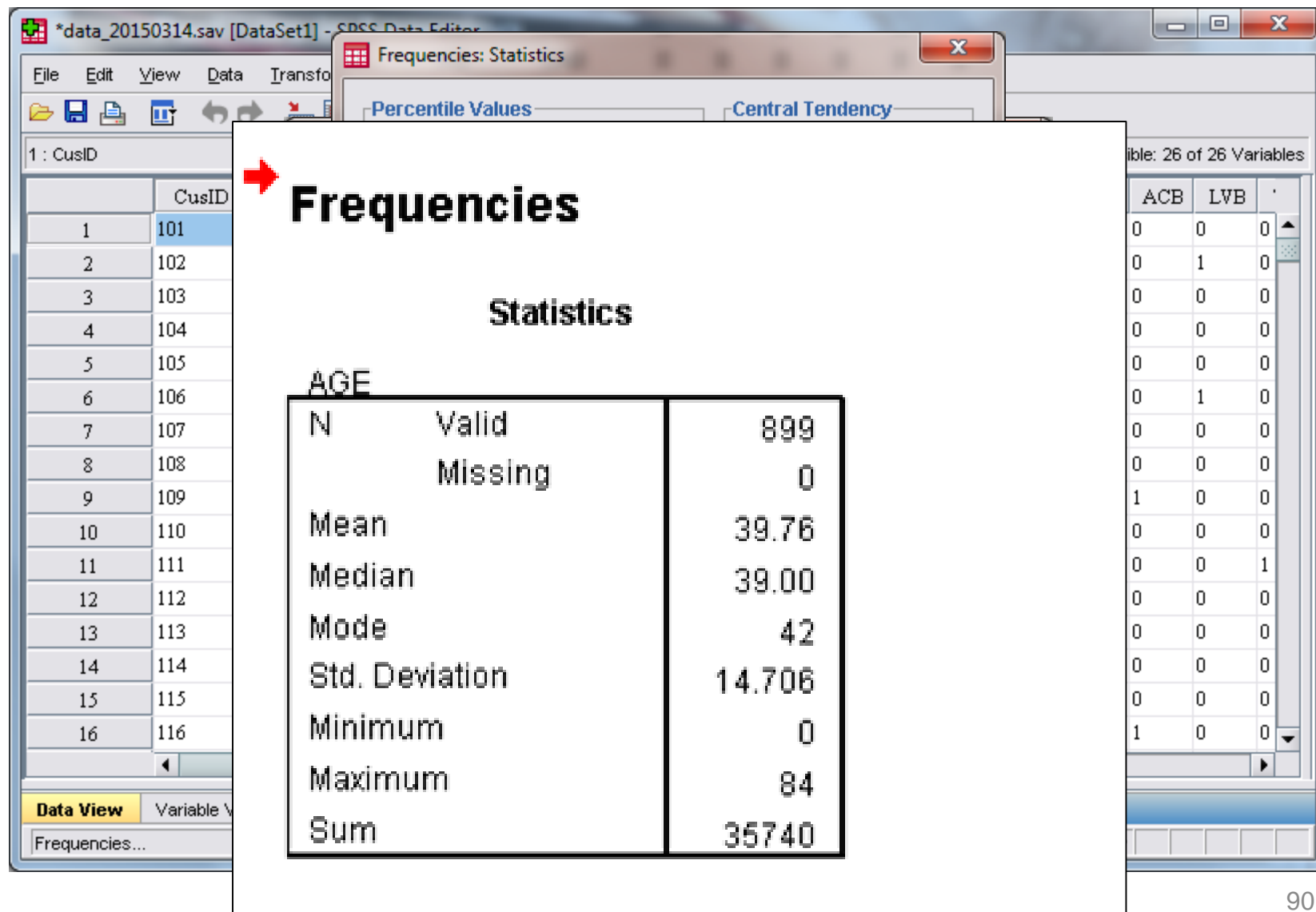
■ Mục đích

- Kết hợp thống kê số lượng đối tượng theo từng biểu hiện của thuộc tính và tính toán thông số thống kê mô tả liên quan đến thuộc tính đó

■ Quy trình thực hiện

- **Analyze** → **Descriptive Statistics** → **Frequencies**
- Chọn biến muốn thống kê trong **Frequencies** → **Statistics** để chọn các thông số thống kê.

V.D: THỐNG KÊ MÔ TẢ (AGE)



Frequencies

Statistics

AGE

N	Valid	899
	Missing	0
Mean		39.76
Median		39.00
Mode		42
Std. Deviation		14.706
Minimum		0
Maximum		84
Sum		35740

90

THỐNG KÊ MÔ TẢ - DESCRIPTIVES

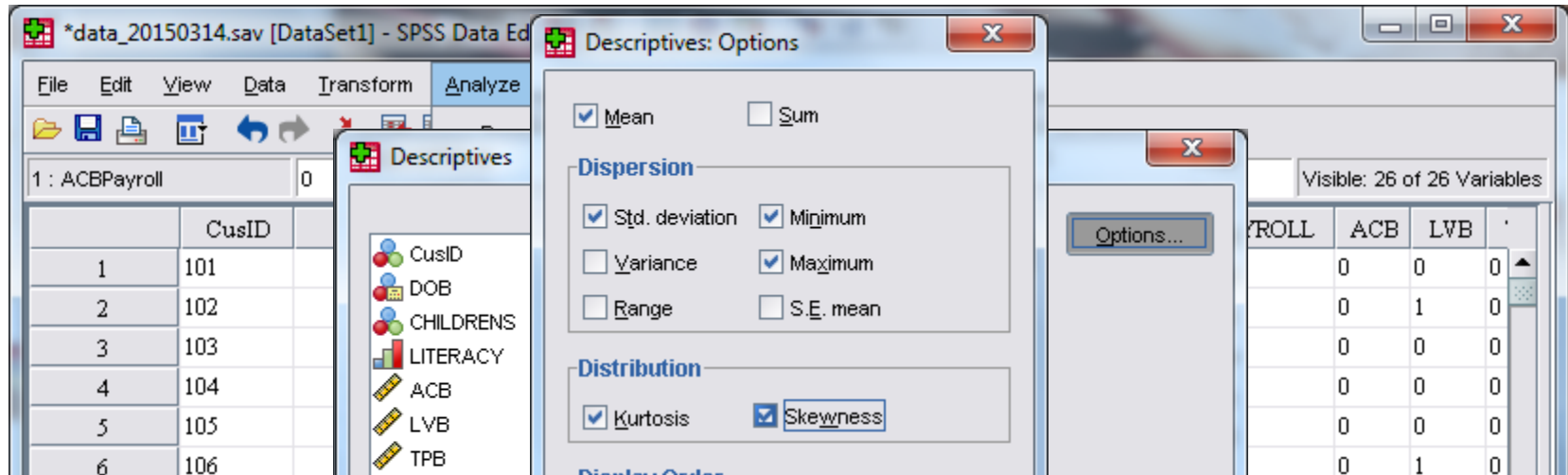
■ Mục đích

- Biểu diễn các biến định lượng của tập dữ liệu nghiên cứu dưới dạng các thông số thống kê mô tả

■ Quy trình thực hiện

- **Analyze** → **Descriptive Statistics** → **Descriptives**
- Chọn biến muốn thống kê trong **Descriptives** → **Options** để chọn các thông số thống kê mô tả.

V.D., THỐNG KÊ MÔ TẢ (AGE)



Descriptives

[DataSet1] D:\02 Tai lieu dao tao\002 Tin hoc ung dung\data_20150314.sav

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
AGE	899	0	84	39.76	14.706	.064	.082	-1.057	.163
Valid N (listwise)	899								

3. THỐNG KÊ MÔ TẢ - EXPLORE

■ Mục đích

- Xác định các thông số thống kê mô tả của một biến trong quan hệ với các biến khác trong bộ dữ liệu

■ Chức năng

- Tính toán các thông số thống kê mô tả trong toàn bộ dữ liệu hoặc theo từng nhóm con của chúng
- Nhận diện các giá trị đặc biệt của bộ dữ liệu
- Tính toán các giá trị thập phân vị của biến trong toàn bộ dữ liệu hoặc trong từng nhóm con của chúng

THỐNG KÊ MÔ TẢ - EXPLORE (TT)

■ Quy trình thực hiện

- **Analyze → Descriptive Statistics → Explore**
- Chọn các biến định lượng muốn so sánh các thông số thống kê mô tả theo nhóm vào khung **Dependent List**
- Chọn các biến định tính muốn sử dụng để phân tách nhóm → **Statistics → Explore: Statistics** và chọn các thống kê tương ứng → **Explore: Plots** và chọn loại biểu đồ thể hiện → **Options** để lựa chọn cách xử lý các giá trị thiếu.

V.D: THỐNG KÊ MÔ TẢ - EXPLORE

The screenshot shows the SPSS Data Editor window with the 'Explore' dialog box open. The 'Explore: Options' sub-dialog is also open, showing the 'Missing values' section. The 'Dependent List' contains 'AGE'. The 'Missing values' section has three radio buttons: 'Exclude cases listwise' (selected), 'Exclude cases pairwise', and 'Report values'. The 'Display' section has three radio buttons: 'Both' (selected), 'Statistics', and 'Plots'. The 'Data View' tab is active, showing a list of variables and their values.

SPSS Data Editor window: *data.sav [DataSet1] - SPSS Data Editor

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

Reports: Descriptive Statistics 123 Frequencies

Visible: 23 of 23 Variables

Explore dialog box:

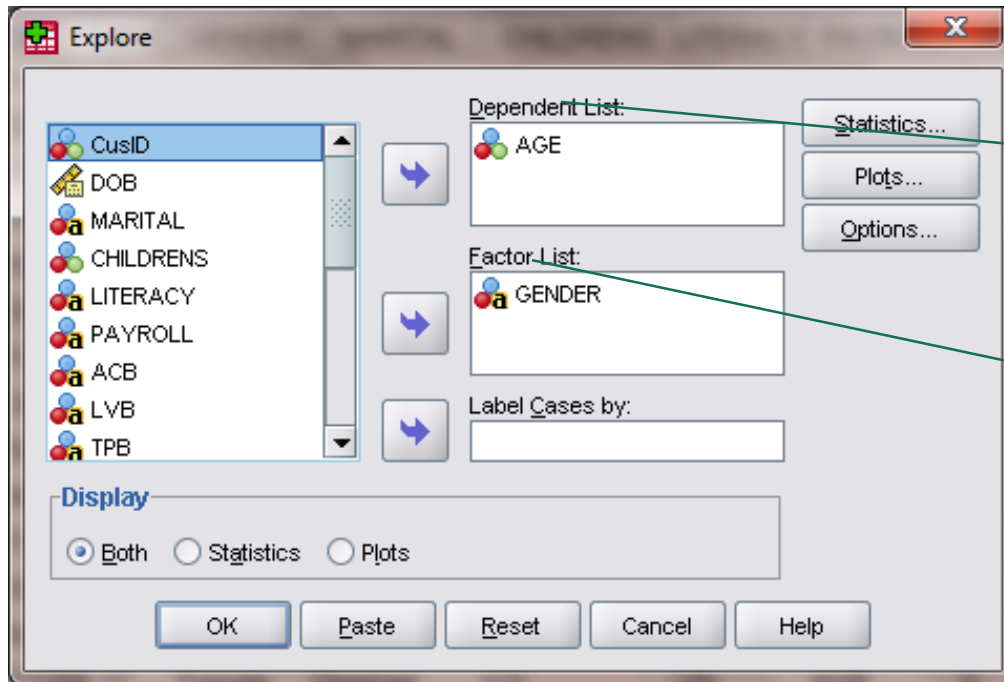
- Dependent List: AGE
- Missing values:
 - ☒ Exclude cases listwise
 - ☐ Exclude cases pairwise
 - ☐ Report values
- Display:
 - ☒ Both
 - ☐ Statistics
 - ☐ Plots

Data View table:

	CusID	DOB	MARITAL	CHILDRENS	LITERACY	PAYROLL	ACB	LVB	TPB	UN	EIB	1	0	1	0	0
1	101	02/														
2	105	12/														
3	109	11/														
4	110	07/														
5	111	05/														
6	112	03/														
7	114	08/														
8	120	04/														
9	121	11/														
10	129	04/														
11	131	09/														
12	132	09/														
13	133	01/														
14	134	05/														
15	143	12/29/1958														

SPSS Processor is ready

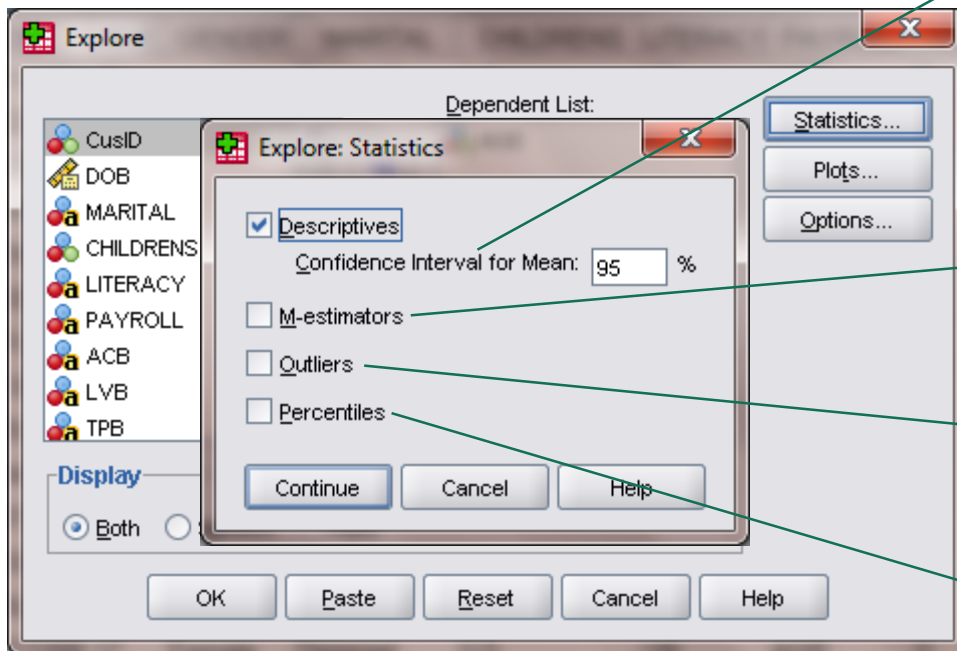
THỐNG KÊ MÔ TẢ EXPLORE



Biến phụ thuộc
(sự khác biệt giữa các nhóm)

Biến nhân tố
(chia dữ liệu thành nhiều nhóm)

THỐNG KÊ MÔ TẢ EXPLORE



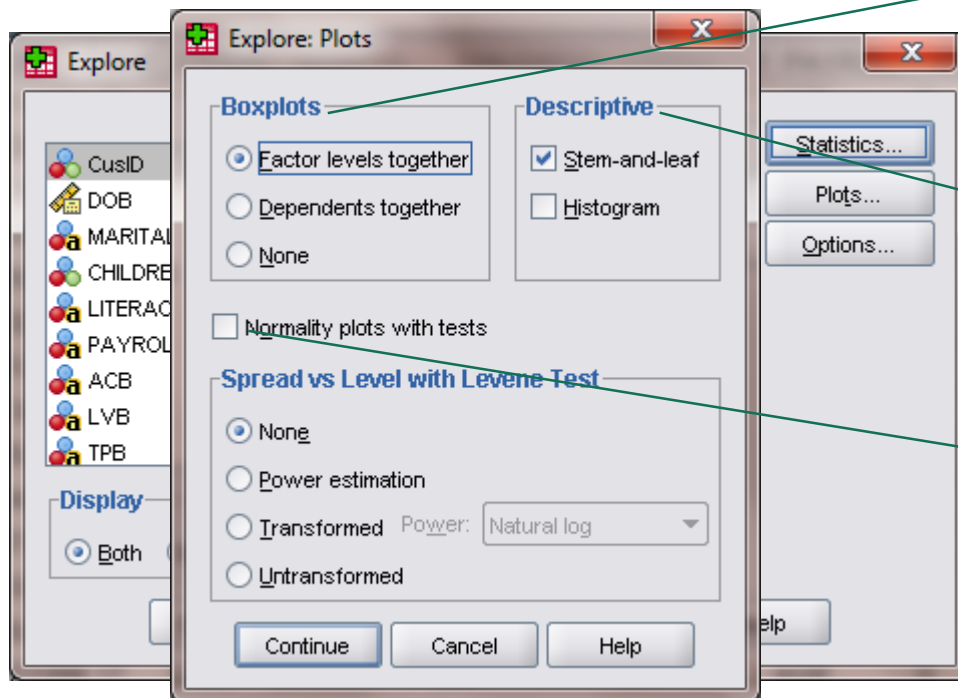
Tính toán các đại lượng thống kê mô tả (độ tin cậy 95%)

Ước lượng các số thống kê tương đồng với số trung bình

Thể hiện 5 giá trị lớn nhất & 5 giá trị nhỏ nhất

Thể hiện các thập phân vị 5, 10, 25, 50, 75, 90, 95

THỐNG KÊ MÔ TẢ EXPLORE

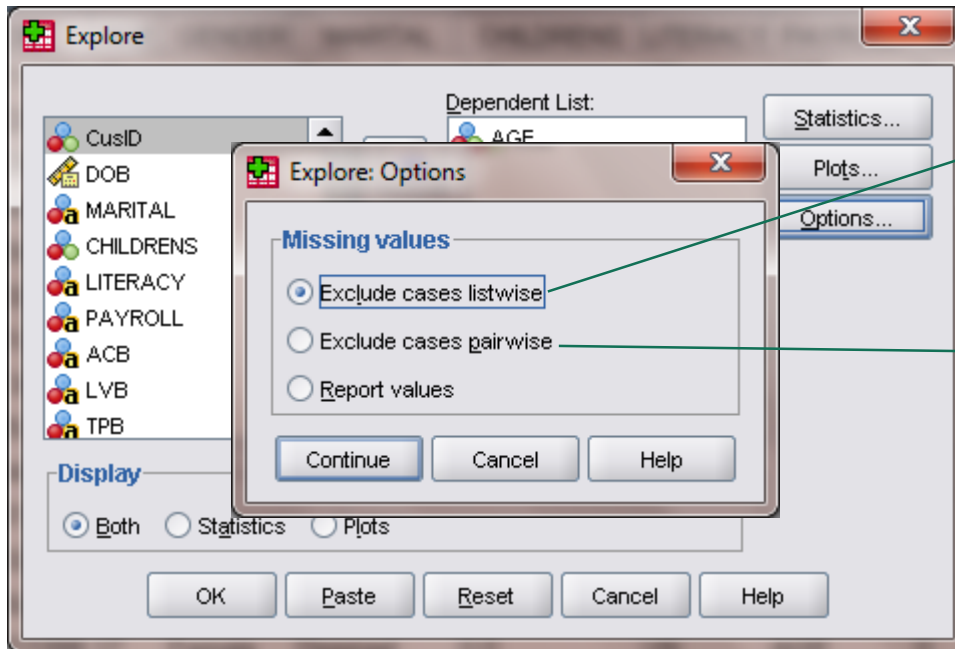


Sắp xếp lại cách thể hiện các biểu đồ hộp Boxplot

Dạng biểu đồ (thân & lá và Histogram)

Vẽ biểu đồ xác suất chuẩn Q-Q plot giúp kiểm tra phân phối chuẩn của biến

THỐNG KÊ MÔ TẢ EXPLORE



Không sử dụng các quan sát thiếu thông tin ở biến phụ thuộc hoặc biến nhân tố

Quan sát không được sử dụng khi tính toán liên quan đến biến bị thiếu thông tin

V.D: THỐNG KÊ MÔ TẢ - EXPLORE (SPV)

→ **Explore**

GENDER

Case Processing Summary

		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
AGE	GENDER	57	100.0%	0	.0%	57	100.0%
	Female	58	100.0%	0	.0%	58	100.0%
	Male	69	100.0%	0	.0%	69	100.0%

V.D: THỐNG KÊ MÔ TẢ - EXPLORE (SPV)

		Skewness	.053	.150
		Kurtosis	-1.200	.300
Female	Mean		41.15	.835
	95% Confidence Interval for Mean	Lower Bound	39.51	
		Upper Bound	42.79	
	5% Trimmed Mean		41.20	
	Median		42.00	
	Variance		222.455	
	Std. Deviation		14.915	
	Minimum		15	
	Maximum		84	
	Range		69	
	Interquartile Range		25	
	Skewness		-.018	.137
	Kurtosis		-.995	.272
	Mean		38.24	.799
Male	95% Confidence Interval for Mean	Lower Bound	36.67	
		Upper Bound	39.81	
	5% Trimmed Mean		38.08	
	Median		37.00	
	Variance		202.143	
	Std. Deviation		14.218	
	Minimum		11	

IBM®

SPSS®

BẢNG KẾT HỢP NHIỀU BIẾN

1. Bảng kết hợp nhiều biến định tính
 - Kết hợp 2 biến định tính
 - Kết hợp 3 biến định tính
2. Bảng kết hợp biến định tính và định lượng
 - Kết hợp 1 biến định tính và 1 biến định lượng
 - Kết hợp 2 biến định tính và 1 biến định lượng.

1. KẾT HỢP NHIỀU BIẾN ĐỊNH TÍNH

- Mục tiêu

- Tính tần số/tần suất các biểu hiện của một biến định tính theo sự phân loại của một (số) biến khác

- Phương pháp

- Dùng Basic/General Tables và/hoặc
- Dùng Custom Tables...

V.D: KẾT HỢP 2 BIẾN ĐỊNH TÍNH

- *Thống kê số lượng Nam và Nữ theo độ tuổi*
- *Thống kê số lượng Nam và Nữ theo từng độ tuổi và tình trạng hôn nhân.*

V.D: KẾT HỢP 2 BIẾN ĐỊNH TÍNH

data.sav [DataSet1] - SPSS Data Editor

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

Crosstabs

1 : CusID

	CusID
1	101
2	102
3	103
4	104
5	105
6	106
7	107
8	108
9	109
10	110
11	111
12	112
13	113
14	114
15	115

ACB
LVB
TPB
EIB
SCB
VCB
VIB
BIDV
DAB
AGR
SACOM
OTHER
PAYMENT
VCB Payroll [V]

☐ Display clustered bar charts
☐ Suppress tabular output

Crosstabs: Statistics

☐ Chi-square
☐ Correlations

Nominal

☐ Contingency coefficient
☐ Phi and Cramer's V
☐ Lambda
☐ Uncertainty coefficient

Ordinal

☐ Gamma
☐ Somers' d
☐ Kendall's tau-b
☐ Kendall's tau-c

Nominal by Interval

☐ Eta

☐ Kappa
☐ Risk
☐ McNemar

☐ Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1

Continue Cancel Help

Exact...
Statistics...
Cells...
Format...

Visible: 23 of 23 Variables

	VB	TPB	EIB	SCB
0	1	1		
0	0	0		
0	0	0		
0	0	0		
0	0	1		
0	0	0		
0	0	0		
0	0	1		
0	0	1		
0	1	1		
1	1	0		
0	0	0		
0	0	0		
0	0	0		
0	0	0		

Data View Variable View

Crosstabs...

Complex Samples
Quality Control
☒ ROC Curve

SPSS Processor is ready

V.D: KẾT HỢP 2 BIẾN ĐỊNH TÍNH (SPV)

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
GENDER * RANGE_OF_AGE	899	100.0%	0	.0%	899	100.0%

GENDER * RANGE_OF_AGE Crosstabulation

Count		RANGE_OF_AGE					Total
		1-18	18-25	25-36	36-45	45-60	
GENDER		18	34	64	46	100	262
	Female	27	35	60	70	128	320
	Male	27	41	84	62	103	317
Total		72	110	208	178	331	899

2. KẾT HỢP BIẾN ĐỊNH TÍNH-ĐỊNH LƯỢNG

- Mục tiêu

- Thông số thống kê mô tả biến định lượng theo sự phân loại của một biến định tính

- Phương pháp

- Dùng **Custom Tables...**

- *V.d: tính tuổi trung bình theo giới tính trong nhóm.*

V.D: KẾT HỢP ĐỊNH TÍNH - ĐỊNH LƯỢNG

Custom Tables

Table Titles Test Statistics Options

Normal Compact Layers

Variables:

Columns

Rows

Categories:

Define

Summary

Position:

Source:

OK

Custom Tables

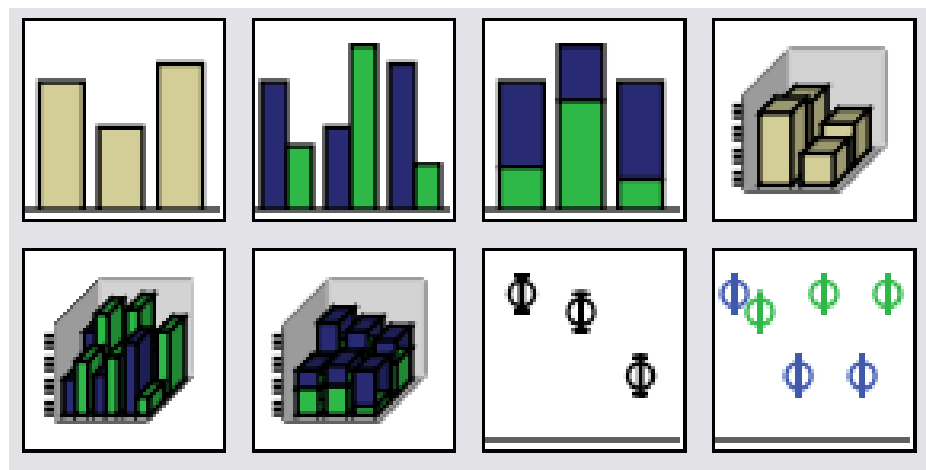
RANGE_OF_AGE	
Mean	
GENDER	
Female	4
Male	4

IBM®

SPSS®

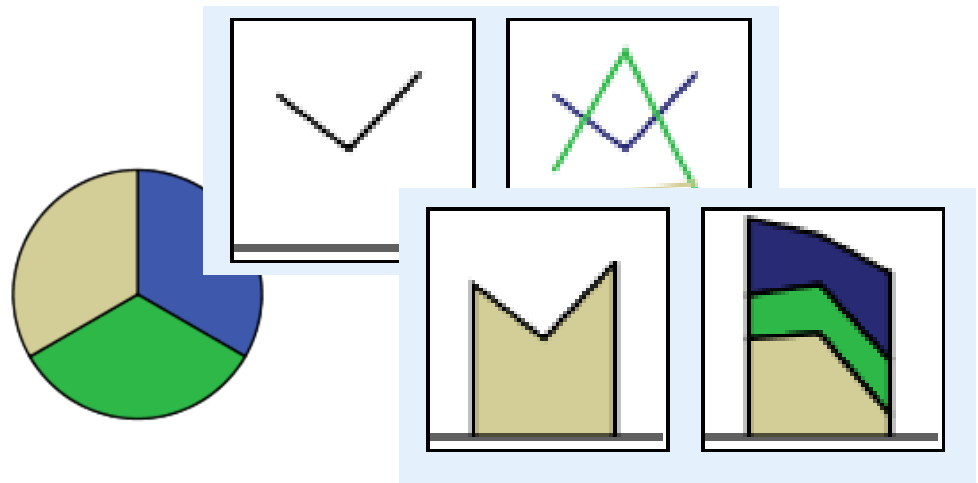
TRÌNH BÀY KẾT QUẢ BẰNG ĐỒ THỊ

1. Các loại đồ thị cơ bản trên SPSS
2. Cách tạo biểu đồ
3. Hiệu chỉnh, bổ sung đồ thị...



1. CÁC LOẠI ĐỒ THỊ TRÊN SPSS

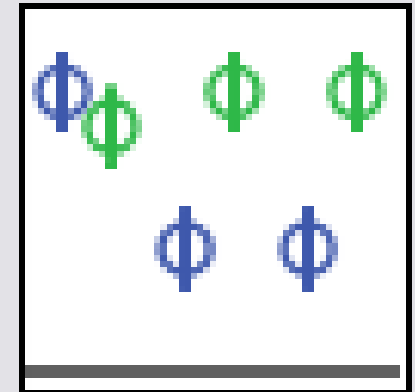
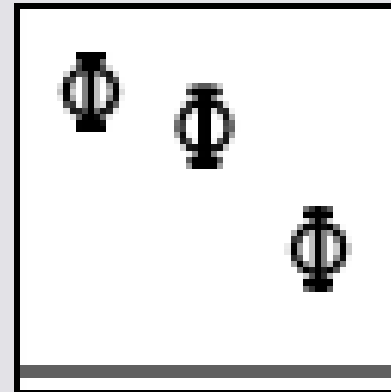
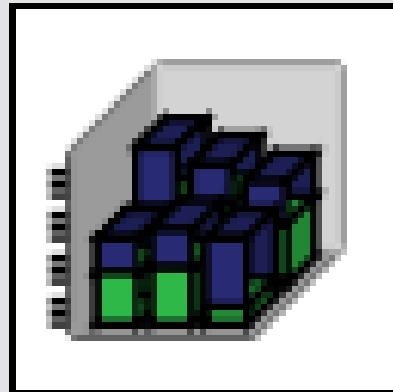
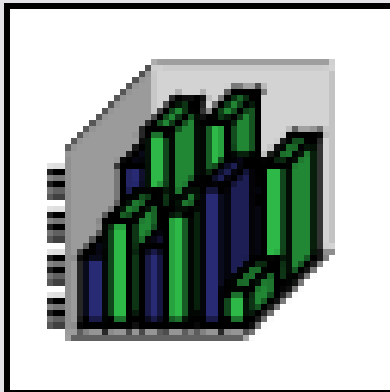
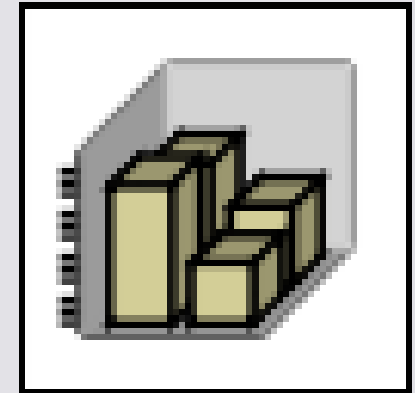
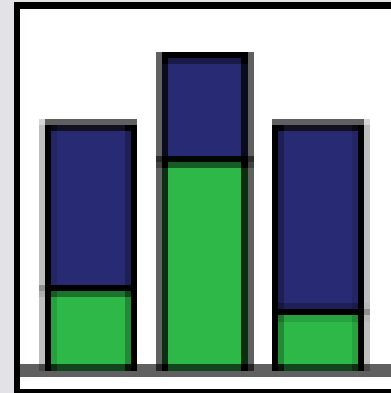
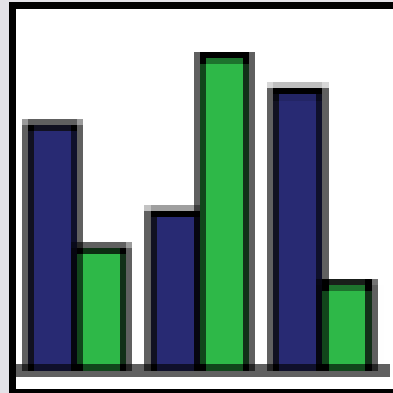
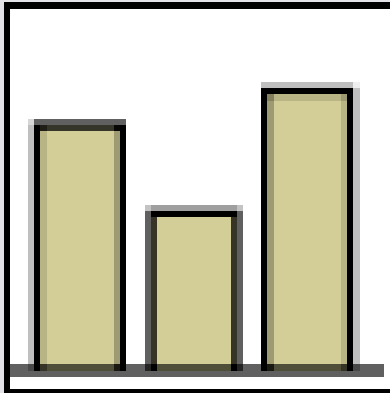
- Biểu đồ thanh, biểu đồ tròn: áp dụng cho dữ liệu định tính
- Biểu đồ hộp và râu, nhật đồ - tần số, biểu đồ tán xạ, biểu đồ gấp khúc... áp dụng cho dữ liệu định lượng.



BIỂU ĐỒ THANH (BAR)

- Biểu đồ thanh được dùng cho dữ liệu định danh hay thứ bậc
- Chiều cao của mỗi thanh biểu diễn tần số hay tần suất của trường hợp biểu diễn bởi thanh đó
- Có thể áp dụng để biểu diễn cho một hay nhiều tập dữ liệu trên cùng một biểu đồ.

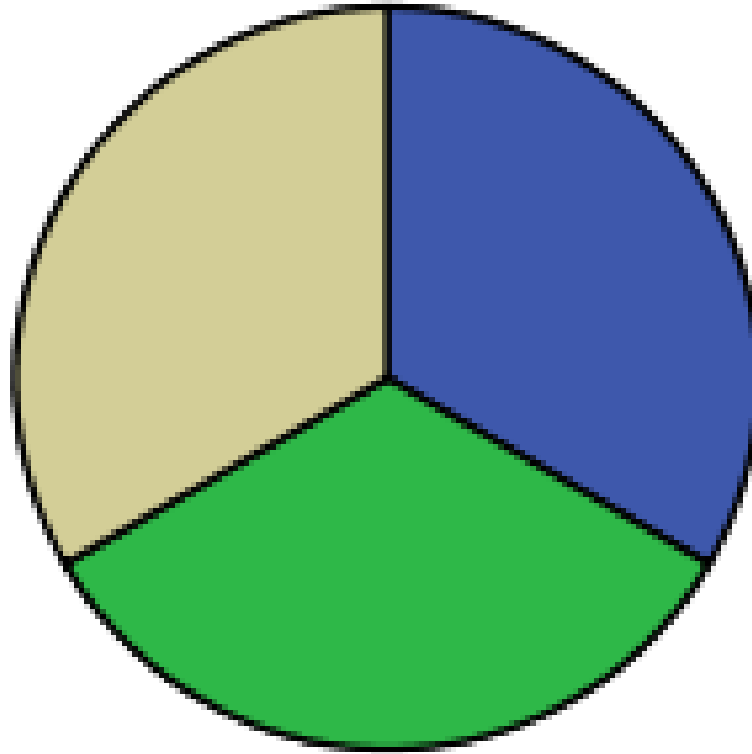
BIỂU ĐỒ THANH (BAR)



BIỂU ĐỒ TRÒN (PIE)

- Biểu đồ tròn được dùng cho dữ liệu định danh hay thứ bậc
- Mỗi hình quạt biểu diễn tỷ lệ đóng góp của từng biểu hiện trong tổng thể hay mẫu.

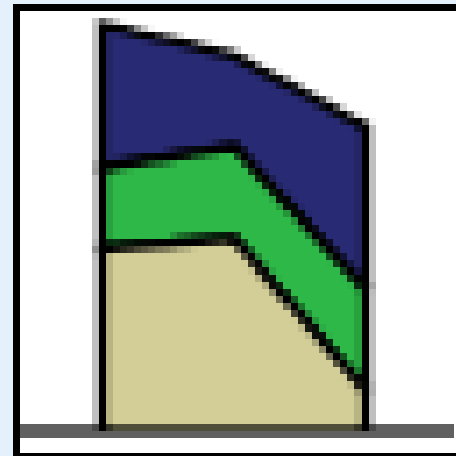
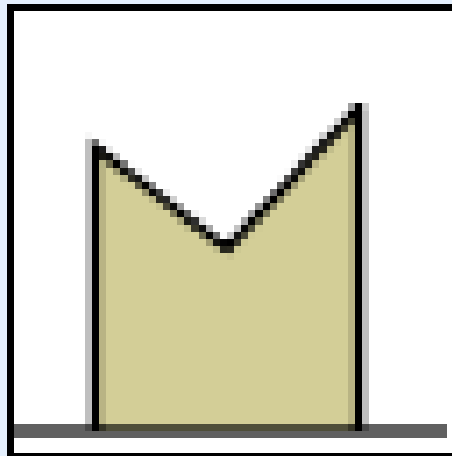
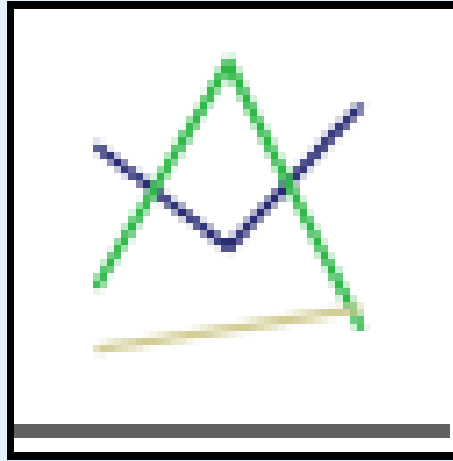
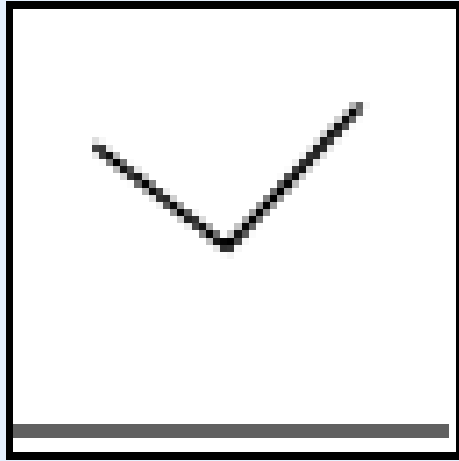
BIỂU ĐỒ TRÒN (PIE)



BIỂU ĐỒ ĐƯỜNG, VÙNG (LINE, AREA)

- Biểu đồ gấp khúc, diện tích dùng để biểu diễn cho những tập dữ liệu định lượng
- Trên một biểu đồ có thể vẽ nhiều đường (vùng diện tích) cùng một lúc
- Hai loại biểu đồ này vẽ giống như biểu đồ thanh.

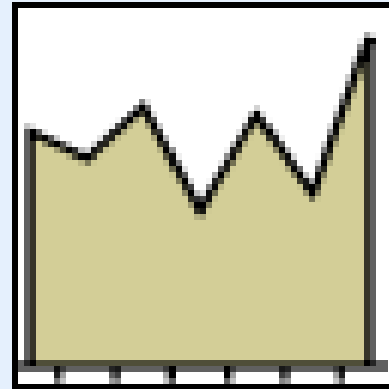
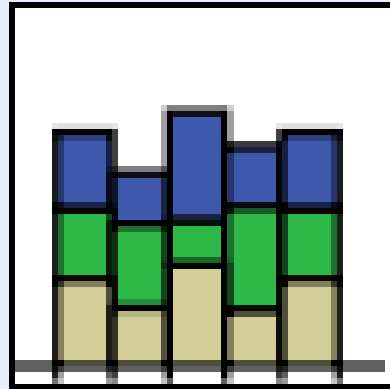
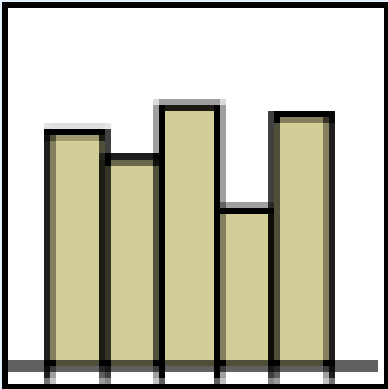
BIỂU ĐỒ ĐƯỜNG, VÙNG (LINE, AREA)



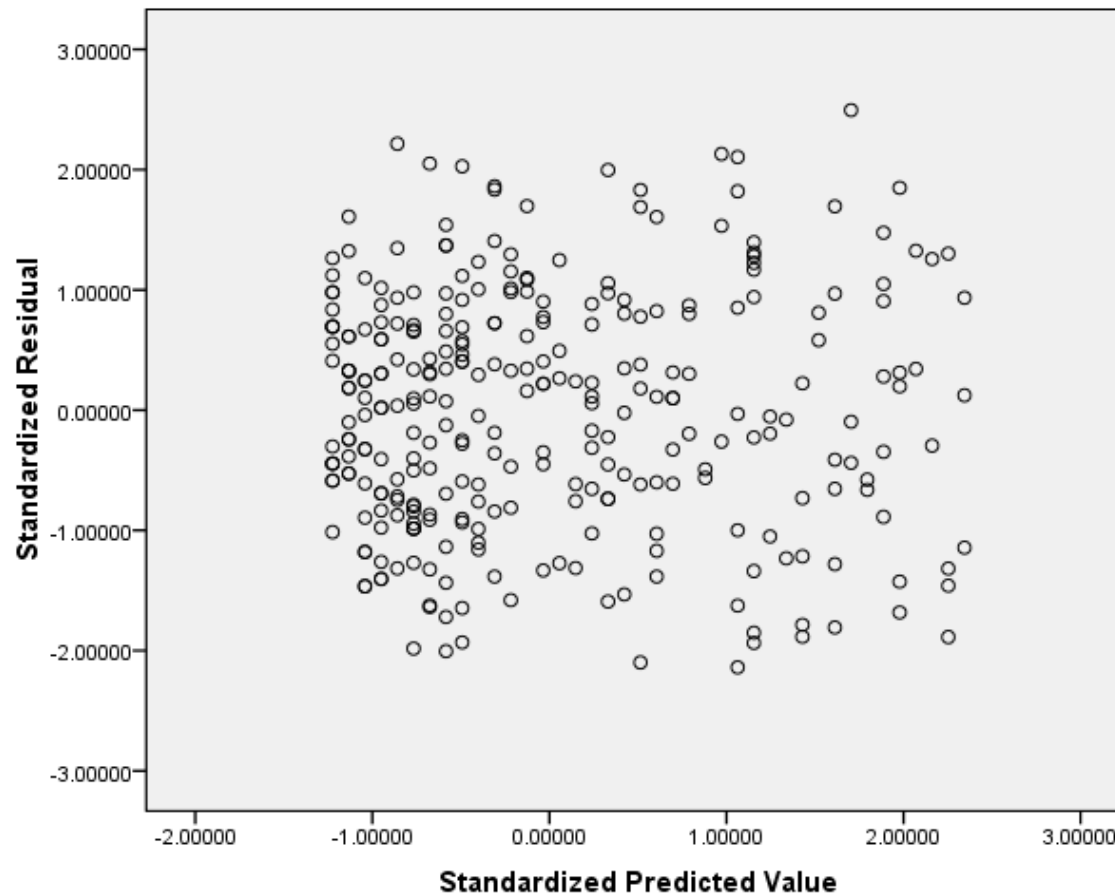
NHẬT ĐỒ - TẦN SỐ (HISTOGRAM)

- Biểu đồ tần số: dùng cho biến định lượng liên tục nhằm biểu diễn phân phối của tập dữ liệu
- Chiều cao mỗi thanh hình chữ nhật trong biểu đồ biểu diễn tần số của lớp tương ứng.

NHẬT ĐỒ - TẦN SỐ (HISTOGRAM)



BIỂU ĐỒ PHÂN TÁN (SCATTER)



2. CÁCH VẼ BIỂU ĐỒ (BAR)

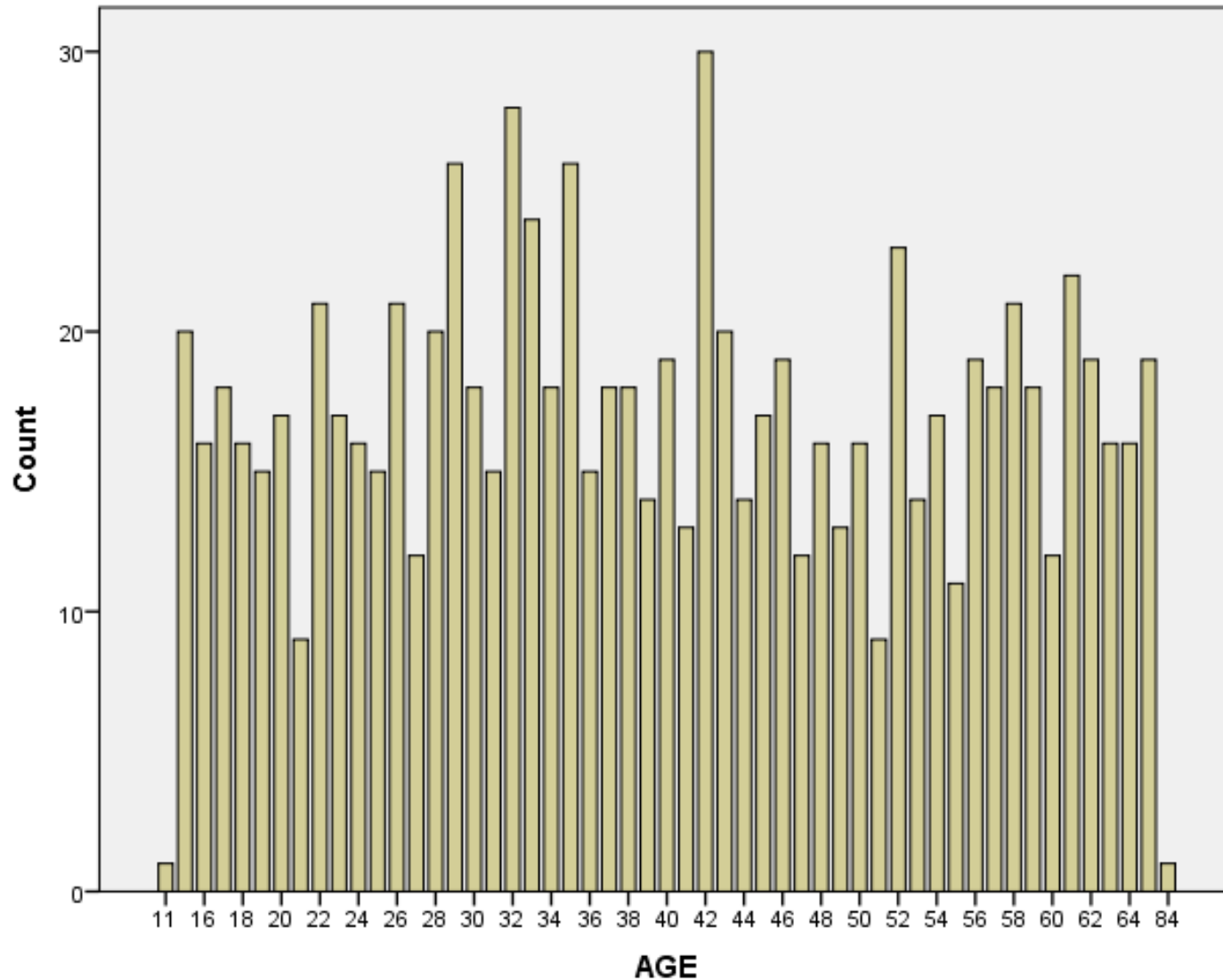
- Cách vẽ biểu đồ thanh:

- **Graph** → **Legacy Dialogs** → **Bar** để mở Bar Chart

- Tại cửa sổ Bar Chart

- **Simple**: nếu biểu diễn tập dữ liệu một biến
- **Clustered**: nếu muốn biểu diễn tập dữ liệu của một biến được chia nhóm bởi một biến khác, các thanh đứng kề nhau
- **Stacked**: nếu muốn biểu diễn tập dữ liệu của một biến được chia nhóm bởi một biến khác, các thanh chồng lên nhau.

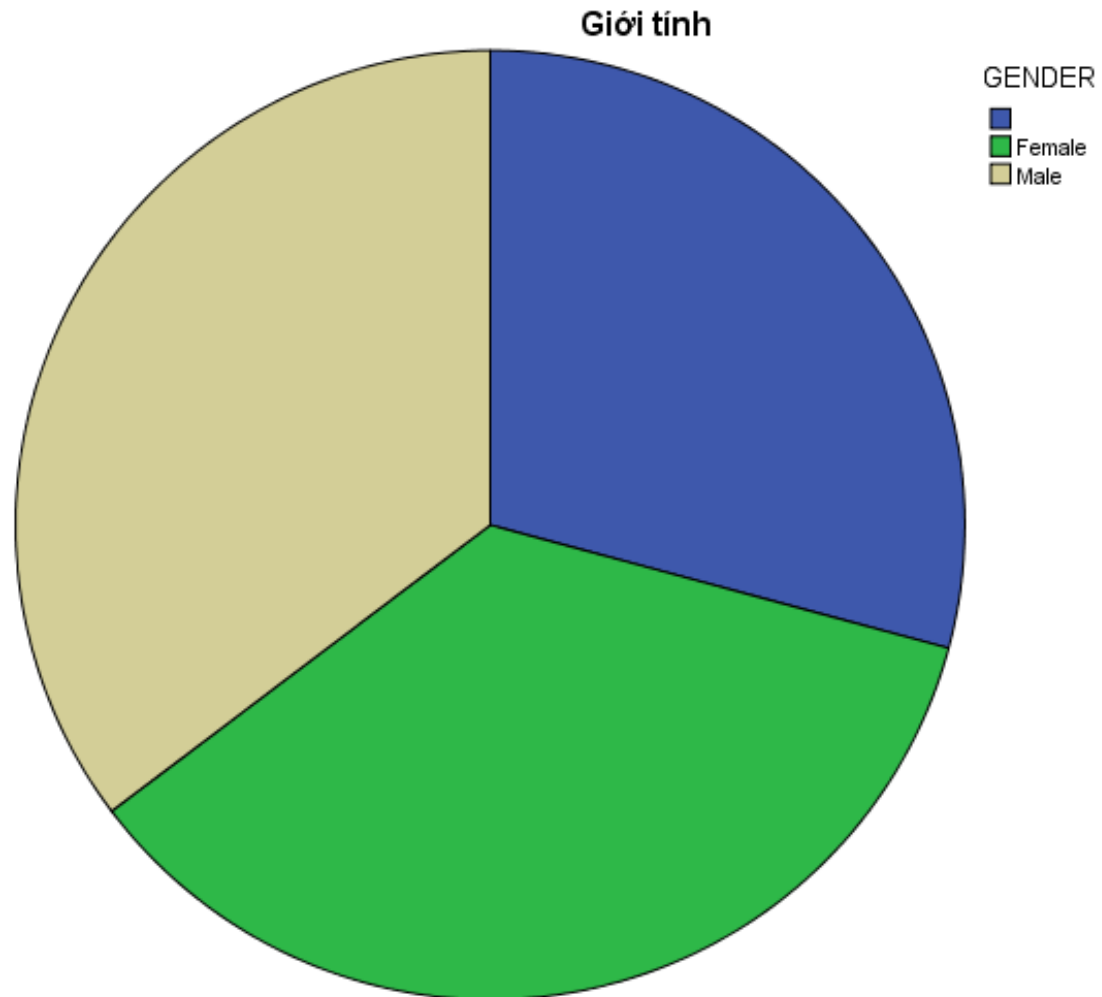
V.D., BIỂU ĐỒ THANH (AGE)



V.D., BIỂU ĐỒ TRÒN (GENDER)

- Pie Chart → Simple...
 - Tại Data in Chart Area → Summary for group of cases → Define → Xuất hiện cửa sổ...
 - Đưa biến Gender vào **Category Axis**
 - Chọn **N of cases** để trực tung biểu diễn tần số, **% of cases** để trực tung biểu diễn tần suất
- **Title** đặt tiêu đề → **OK**...

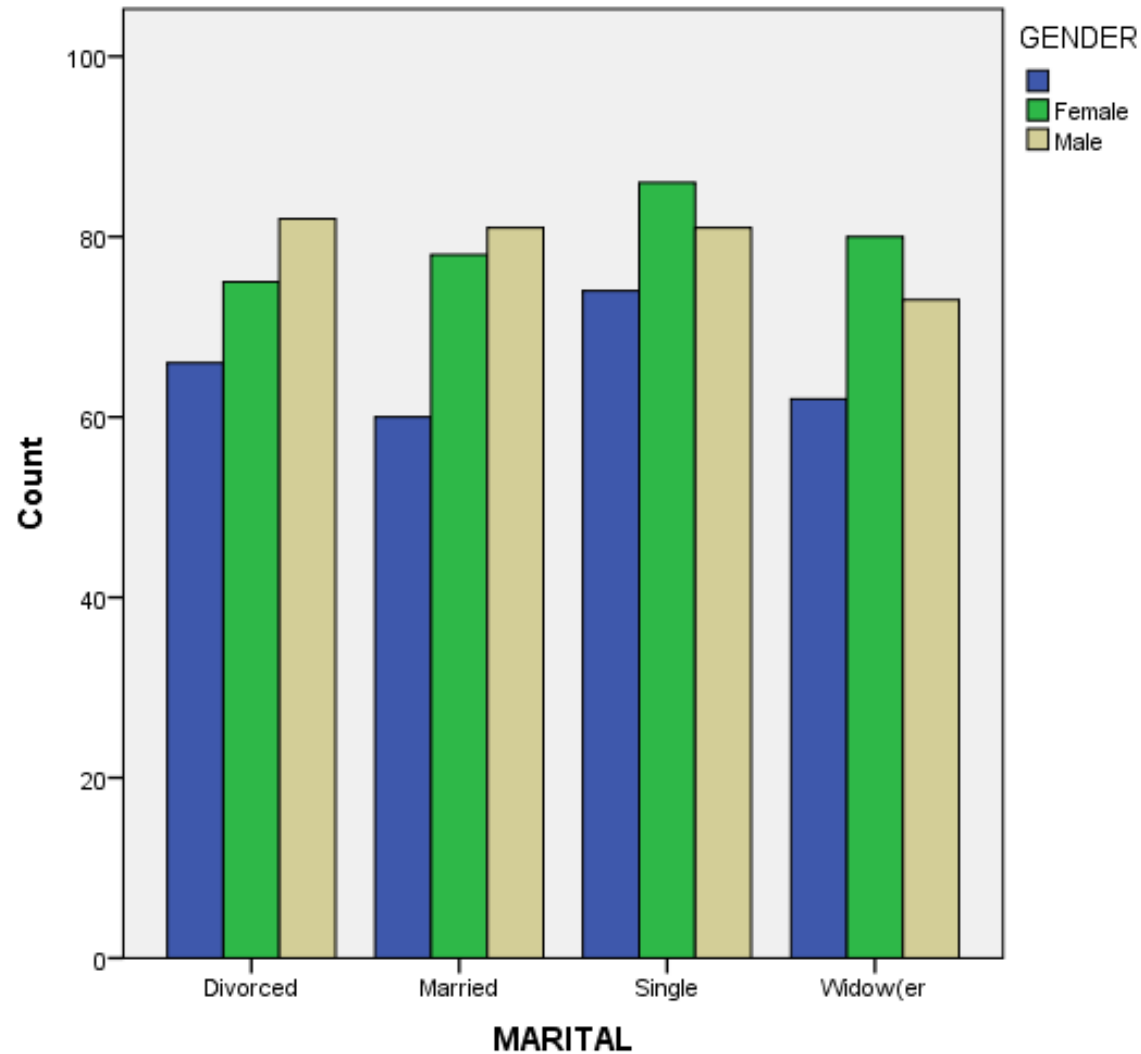
V.D., BIỂU ĐỒ TRÒN (GENDER)



BIẾN ĐƯỢC PHÂN TÍCH BỞI BIẾN KHÁC

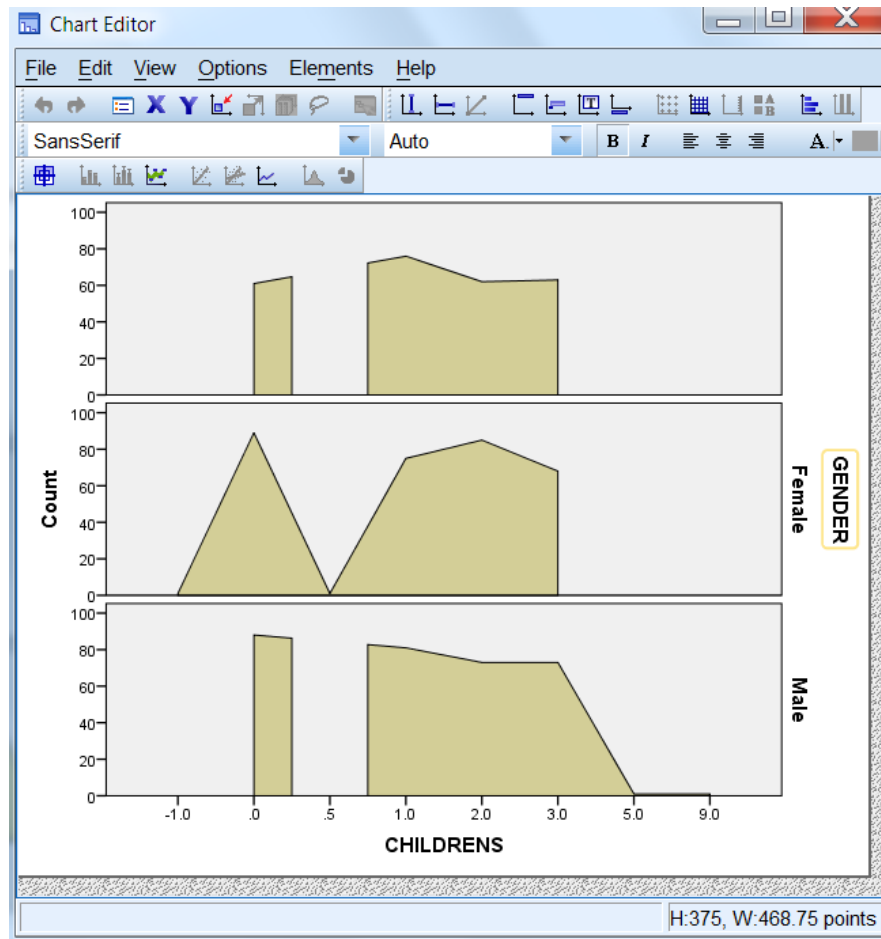
- Bar Chart → Clustered...
 - Tại Data in Chart Area → Summary for group of cases → Define...
 - Đưa biến *Marital* vào **Category Axis**, đưa biến *Gender* vào khung **Define Cluster by**
 - Chọn **N of cases** để trực tung biểu diễn tần số, **% of cases** để trực tung biểu diễn tần suất
- **Title** đặt tiêu đề → **OK...**

V.D: BIẾN ĐƯỢC PHÂN TÍCH BỞI BIẾN KHÁC



3. HIỆU CHỈNH BIỂU ĐỒ

- Trong Viewer, nhấp chuột 2 lần vào đồ thị và sửa...



HIỆU CHỈNH BIỂU ĐỒ (TT)

- Cửa sổ **Chart Editor** xuất hiện
- Ở cửa sổ này sẽ thực hiện các thao tác **Edit** biểu đồ của mình như màu sắc, đưa số liệu hiển thị lên biểu đồ, chú thích cho biểu đồ
- Có thể thay đổi màu, chọn loại phong hoặc cỡ chữ, chuyển đổi trực tung với trục hoành, xoay các đồ thị
- Thậm chí thay cả loại đồ thị...

IBM®

SPSS®

MỐI LIÊN HỆ GIỮA 2 BIẾN ĐỊNH TÍNH

1. Kiểm định mối liên hệ giữa 2 biến danh nghĩa
2. Mối liên hệ giữa biến danh nghĩa và thứ bậc
3. Kiểm định mối liên hệ giữa 2 biến thức bậc

[Thang đo >>](#)

1/2. MỐI LIÊN HỆ GIỮA 2 BIẾN [DN/TB]

Kiểm định mối liên hệ giữa 2 biến định tính [dn/tb]

- Giả thuyết
 - H_0 : Không tồn tại mối liên hệ giữa 2 biến
 - H_1 : Hai biến có mối liên hệ với nhau
- Phương pháp
 - Sử dụng kiểm định *Chi-square* (χ^2)
 - Mức ý nghĩa $\alpha = 0.05$ (95%)

KIỂM ĐỊNH *CHI-SQUARE* (χ^2)

- Kiểm định *Chi-square* (χ^2) được sử dụng trong việc kiểm định mối liên hệ giữa 2 biến
 - Danh nghĩa - danh nghĩa
 - Danh nghĩa - thứ bậc
- Cho biết có tồn tại hay không mối liên hệ giữa 2 biến trong tổng thể.

χ^2

KIỂM ĐỊNH *CHI-SQUARE* (χ^2)

- Bước 1: Giả thuyết H_0 :
“*Không có mối liên hệ giữa hai biến*”
- Bước 2: Thực hiện kiểm định *Chi-square* (χ^2)
- Bước 3: So sánh *p-value* với giá trị α
 - *p-value* $> \alpha$: Chấp nhận H_0
 - *p-value* $\leq \alpha$: Bác bỏ H_0

KIỂM ĐỊNH *CHI-SQUARE* (χ^2)

- Trong SPSS, *p-value* là giá trị *Sig.*
- α là khả năng tối đa bác bỏ H_0
 - Nếu $\alpha = 5\%$ thì chấp nhận khả năng sai lầm tối đa là 5%
 - Độ tin cậy của kiểm định là $1 - \alpha = 95\%$

Fisher (1925)

KIỂM ĐỊNH *CHI-SQUARE* (χ^2)

- Thực hiện:
 - **Analyze** → **Descriptive Statistics** → **Crosstabs**
 - Chọn các biến vào **Row(s)** và **Column(s)**
 - **Statistics** → chọn **Chi-square** → **OK...**

V.D., MỐI LIÊN HỆ GIỮA 2 BIẾN DANH NGHĨA

data.sav [DataSet1] - SPSS Data Editor

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

Reports Descriptive Statistics 123 Frequencies...

1 : CusID

Visible: 23 of 23 Variables

Crosstabs

Row(s):
a GENDER

Column(s):
a LITERACY

Layer 1 of 1

Previous Next

☐ Display clustered bar charts
☐ Suppress tables

OK Paste Reset Cancel Help

Exact... Statistics... Cells... Format...

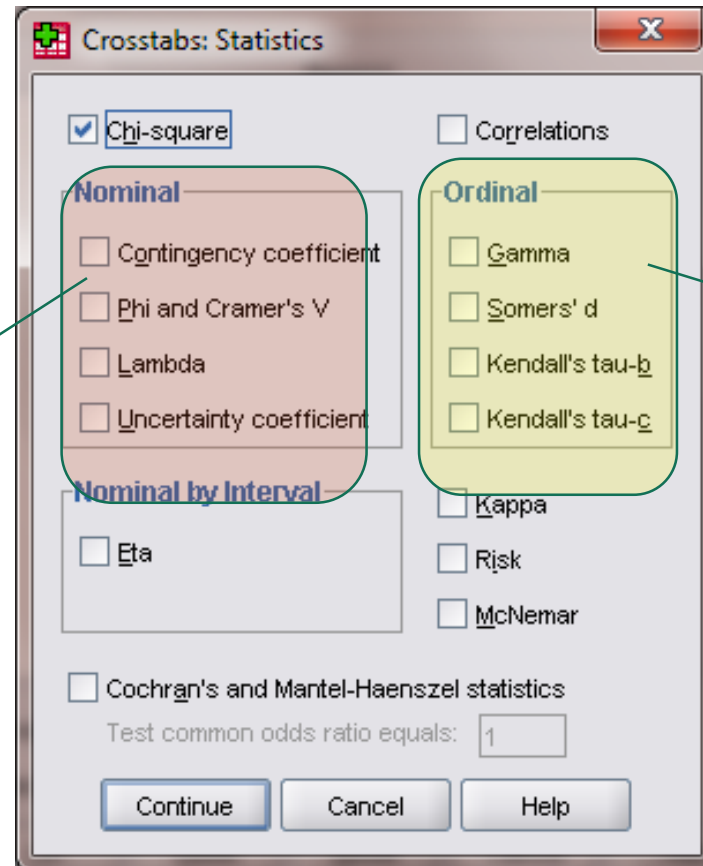
	VB	TPB	EIB	SCB
0	1	1		
0	0	0		
0	0	0		
0	0	0		
0	0	1		
0	0	0		
0	0	0		
0	0	1		
0	0	1		
0	1	1		
1	1	0		
0	0	0		
0	0	0		
0	0	0		

Data View Variable View

Crosstabs...

SPSS Processor is ready

V.D., MỐI LIÊN HỆ GIỮA 2 BIẾN DANH NGHĨA



Các đại lượng kiểm định dành cho 2 biến danh nghĩa

Các đại lượng kiểm định dành cho 2 biến thứ bậc

MỐI LIÊN HỆ GIỮA 2 BIẾN DANH NGHĨA (SPV)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
GENDER * LITERACY	899	100.0%	0	.0%	899	100.0%

GENDER * LITERACY Crosstabulation

Count		LITERACY						Total
		CO	GE	HS	LS	PS	UN	
GENDER								
	Female	44	16	15	4	6	177	262
	Male	44	24	17	10	11	214	320
		41	28	21	6	12	209	317
Total		129	68	53	20	29	600	899

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.527 ^a	10	.769
Likelihood Ratio	6.512	10	.771
N of Valid Cases	899		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.83.

Do *Sig.* > α nên chấp nhận giả thuyết H_0

3. MỐI LIÊN HỆ GIỮA 2 BIẾN [TB]

Kiểm định mối liên hệ giữa 2 biến định tính [tb]

- Giả thuyết

- H_0 : Không tồn tại mối liên hệ giữa 2 biến
- H_1 : Hai biến có mối liên hệ với nhau

- Phương pháp

- Sử dụng các kiểm định tau-b (*Kendall*), d (*Some*), gamma (*Goodman & Kruskal*)
- Mức ý nghĩa $\alpha = 0.05$ (95%)

V.D: MỐI LIÊN HỆ GIỮA 2 BIẾN THỨ BẬC

data.sav [DataSet1] - SPSS Data Editor

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

1 : CusID

	CusID
1	101
2	102
3	103
4	104
5	105
6	106
7	107
8	108
9	109
10	110
11	111
12	112
13	113
14	114
15	115

Crosstabs

Crosstabs: Statistics

☒ Chi-square ☐ Correlations

Nominal

- ☐ Contingency coefficient
- ☐ Phi and Cramer's V
- ☐ Lambda
- ☐ Uncertainty coefficient

Ordinal

- ☐ Gamma
- ☐ Somers' d
- ☐ Kendall's tau-b
- ☐ Kendall's tau-c

Nominal by Interval

- ☐ Eta

☐ Kappa

☐ Risk

☐ McNemar

☐ Cochran's and Mantel-Haenszel statistics

Test common odds ratio equals: 1

Continue Cancel Help

Exact... Statistics... Cells... Format...

Visible: 23 of 23 Variables

	LVB	TPB	EIB	SCB
0	0	1	1	
1	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	1	
1	0	0	0	
0	0	0	0	
0	0	0	1	
0	0	0	1	
0	0	1	1	
0	1	1	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	

Data View Variable View

Crosstabs...

Complex Samples

Quality Control

ROC Curve

SPSS Processor is ready

MỐI LIÊN HỆ GIỮA 2 BIẾN THỨ BẬC (SPV)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15.261 ^a	12	.227
Likelihood Ratio	15.682	12	.206
Linear-by-Linear Association	3.035	1	.081
N of Valid Cases	899		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 16.42.

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.049	.027	1.811	.070
	Gamma	.066	.036	1.811	.070
N of Valid Cases		899			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Do $Sig. > \alpha$ nên chấp nhận giả thuyết H_0

IBM®

SPSS®

KIỂM ĐỊNH TRUNG BÌNH TỔNG THỂ

1. So sánh trị trung bình của 1 tổng thể
2. So sánh trị trung bình của 2 nhóm tổng thể
3. So sánh trị trung bình nhiều nhóm tổng thể

KIỂM ĐỊNH TRUNG BÌNH TỔNG THỂ

- Một số yêu cầu về dữ liệu:
 - Các biến liên tục (thang đo khoảng/tỷ lệ)
 - Các quan sát độc lập
 - Các biến có phân phối chuẩn
 - Phương sai đồng nhất
 - Không có trường hợp bất thường...

1. KIỂM ĐỊNH TRUNG BÌNH 1 TỔNG THỂ

- Bước 1: Giả thuyết H_0 :

“Giá trị trung bình của tổng thể bằng giá trị cho trước”

- Bước 2: Lọc các trường hợp thỏa điều kiện (nếu có)...
- Bước 3: Thực hiện kiểm định *One-Sample T-Test*
- Bước 4: So sánh *p-value (Sig.)* với giá trị α
 - *Sig. > α* : Chấp nhận H_0
 - *Sig. $\leq \alpha$* : Bác bỏ H_0

One-Sample T-Test

1. KIỂM ĐỊNH TRUNG BÌNH 1 TỔNG THỂ

- Ví dụ: Kiểm định giả thuyết “Độ tuổi trung bình của khách hàng trả lương qua thẻ VCB là 35”
- Quy trình thực hiện
 - **Select Case** lọc các trường hợp trả lương qua thẻ VCB...
 - **Analyze → Compare Means → One-Sample T-Test**
 - Cung cấp biến kiểm định, giá trị trung bình kỳ vọng
 - Chọn độ tin cậy...

V.D., KIỂM ĐỊNH TRUNG BÌNH 1 TỔNG THỂ

Giả thuyết H_0 : Tuổi trung bình của chủ thẻ lương VCB là 35

The screenshot shows the SPSS Data Editor window with a dataset named 'data.sav [DataSet1]'. The 'Data View' tab is active, displaying a table with columns: CusID, DOB, CHILDREN, PAYMENT, RANGE_C, VCB Payr, and VCBPayr. The 'One-Sample T Test' dialog box is open, and the 'One-Sample T Test: Options' sub-dialog box is also open. The 'Options' dialog shows a 'Confidence Interval' of 99% and 'Missing Values' set to 'Exclude cases analysis by analysis'. The 'Continue' button in the 'Options' dialog is highlighted with a red rectangle.

	CusID	DOB	CHILDREN	PAYMENT	RANGE_C	VCB Payr	VCBPayr
1	101	02/11/1993					
2	140	02/11/1993					
3	145	03/11/1993					
4	158	08/11/1993					
5	167	09/11/1993					
6	169	11/11/1993					
7	172	07/11/1993					
8	180	12/11/1993					
9	214	11/11/1993					
10	224	03/11/1993					
11	226	08/11/1993					
12	238	07/11/1993					
13	255	02/11/1993					
14	288	11/29/1993					
15	326	11/27/1993					

KIỂM ĐỊNH TRUNG BÌNH 1 TỔNG THỂ (SPV)

Giả thuyết H_0 : Tuổi trung bình của chủ thẻ lương VCB là 35

Số lượng quan sát

Giá trị trung bình

[DataSet1] E:\01 Banking University\05 Tai lieu giang day\01 Tin hoc ung d

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
AGE	62	41.79	13.339	1.694

Mức ý nghĩa = $0.00 < 0.05$

Giá trị kiểm định t

Test Value = 35				
	t	df	Sig. (2-tailed)	Mean Difference
AGE	4.008	61	.000	6.79

Bác bỏ giả thuyết H_0

KẾT LUẬN: Tuổi trung bình của chủ thẻ lương VCB trên 35

2. KIỂM ĐỊNH TRUNG BÌNH 2 TỔNG THỂ

- a. Mẫu độc lập (Independent)
- b. Mẫu phụ thuộc hoặc theo từng cặp (Paired)

TRUNG BÌNH CỦA 2 TỔNG THỂ - ĐỘC LẬP

- Bước 1: Giả thuyết H_0 :

“Giá trị trung bình của 2 biến tổng thể là như nhau”

- Bước 2: Thực hiện *Independent-Samples T-Test*

Independent-Sample T-Test

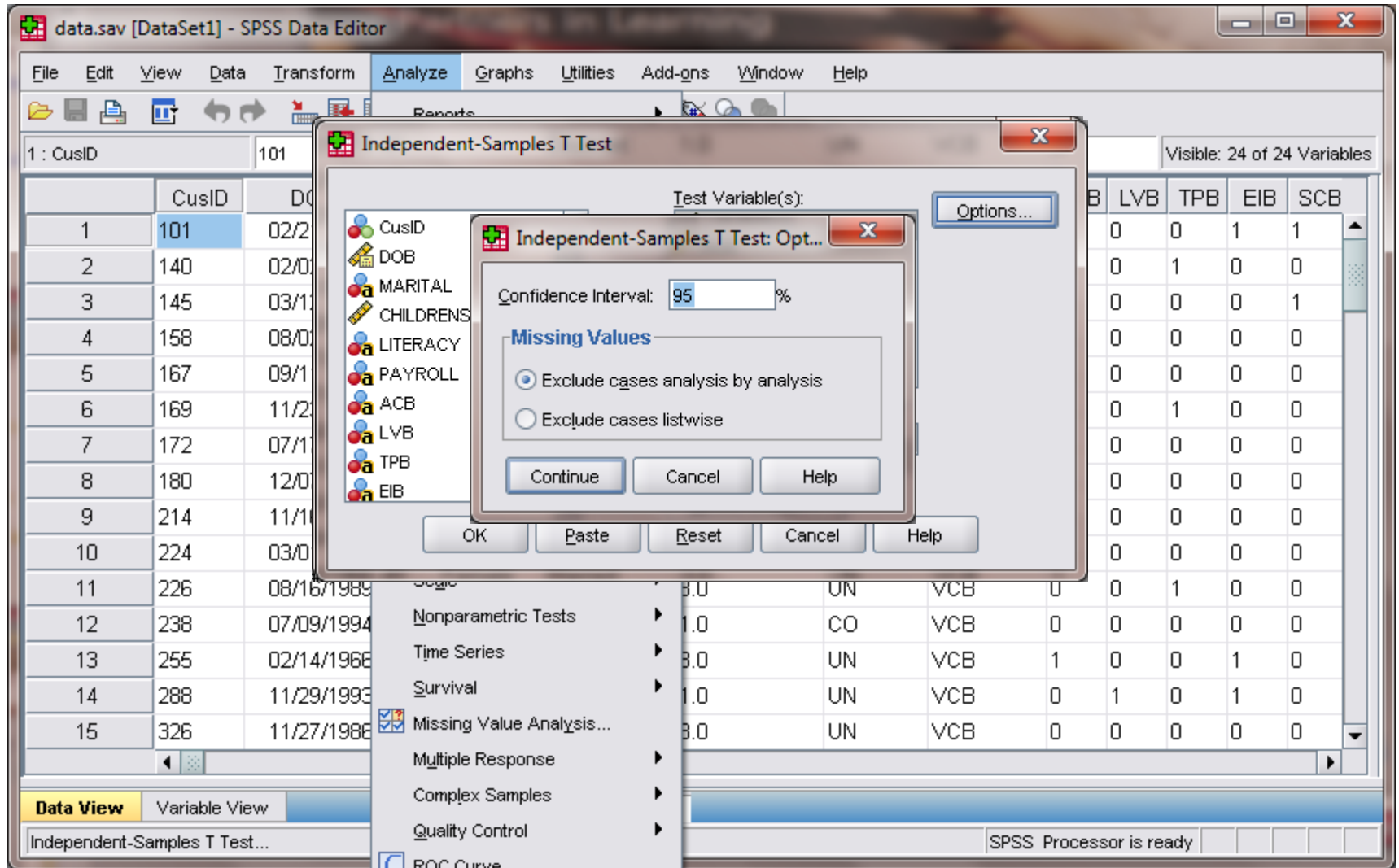
TRUNG BÌNH CỦA 2 TỔNG THỂ - ĐỘC LẬP

- Bước 3: Tìm *Sig.* với kiểm định sự bằng nhau của 2 phương sai tổng thể Levene:
 - Nếu *Sig.* < α thì phương sai giữa 2 nhóm là khác nhau
→ dùng kết quả **“Equal variances not assumed”**
 - Nếu *Sig.* $\geq \alpha$ thì phương sai giữa 2 nhóm là bằng nhau
→ dùng kết quả **“Equal variances assumed”**
- Bước 4: So sánh *Sig.* của kiểm định t ở bước 3 với α :
 - *Sig.* > α : Chấp nhận H_0
 - *Sig.* $\leq \alpha$: Bác bỏ H_0

TRUNG BÌNH CỦA 2 TỔNG THỂ - ĐỘC LẬP

- Ví dụ: Kiểm định giả thuyết “Tuổi trung bình của khách hàng nam và nữ là ngang nhau”
- Quy trình thực hiện
 - **Analyse → Compare Means → Independent-samples T Test**
 - Cung cấp các biến định lượng (tính trung bình), biến định tính (phân thành 2 nhóm độc lập)
 - Chỉ định 2 nhóm cần so sánh...

V.D., TRUNG BÌNH CỦA 2 TỔNG THỂ - ĐỘC LẬP



TRUNG BÌNH CỦA 2 TỔNG THỂ - ĐỘC LẬP (SPV)

Giả thuyết H_0 : Tuổi trung bình giữa nam và nữ là bằng nhau

Kiểm định sự bằng nhau về
phương sai của 2 mẫu
(Levenve's Test)

$Sig. < 5\% \rightarrow$ phương sai 2 mẫu
khác nhau
"Equal variances not assumed"

Independent Samples Test			
		AGE	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	1.041	
	Sig.	.308	
t-test for Equality of Means	t	-2.399	-2.400
	df	635	633.503
	Sig. (2-t)		
	Mean D		
	Std. Err		
	95% Co of the D		

$Sig. > 5\% \rightarrow$ phương sai 2 mẫu
bằng nhau
"Equal variances assumed"

TRUNG BÌNH CỦA 2 TỔNG THỂ - ĐỘC LẬP (SPV)

Kết quả kiểm định sự
bằng nhau về trị trung
bình của 2 mẫu

$Sig. > 5\% \rightarrow$ phương sai 2
mẫu bằng nhau
"Equal variances assumed"

Independent Samples Test			
		ACE	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	1.041	
		.308	
t-test for Equality of Means	t	-2.399	-2.400
	df	635	633.503
	Sig. (2-tailed)	.017	.017
	Mean Difference	-2.785	-2.785
	Std. Error Difference	1.161	1.161
	95% Confidence Interval of the Difference	-5.065	-5.065
		-5.065	-5.065

Mức ý nghĩa = $0.00 < 0.05$

Bác bỏ giả thuyết H_0

KẾT LUẬN: Tuổi trung bình giữa nam và nữ là khác nhau

TRUNG BÌNH CỦA 2 TỔNG THỂ - TỪNG CẶP

- Bước 1: Giả thuyết H_0 :

“Không có sự khác nhau về trị 2 trung bình tổng thể”

(Khác biệt giữa 2 trung bình là bằng 0)

- Bước 2: Thực hiện *Paired-Samples T-Test*

- Bước 3: So sánh *Sig.* của kiểm định t ở bước 2 với α :

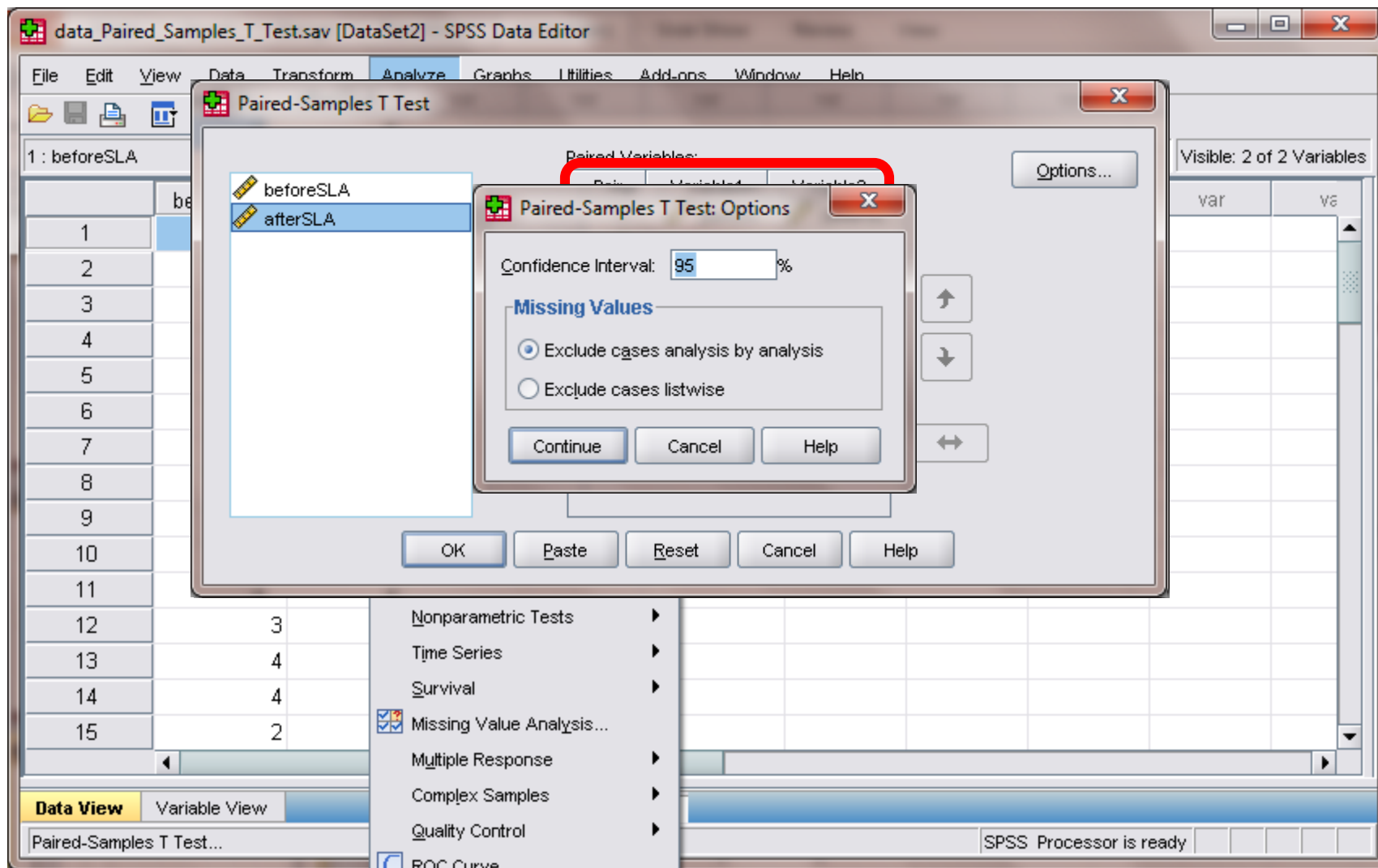
- *Sig.* $> \alpha$: Chấp nhận H_0
- *Sig.* $\leq \alpha$: Bác bỏ H_0

Paired-Sample T-Test

TRUNG BÌNH CỦA 2 TỔNG THỂ - TỪNG CẶP

- Ví dụ: Cảm nhận của khách hàng trước và sau khi ngân hàng triển khai cam kết chất lượng dịch vụ (SLA)
- Quy trình thực hiện
 - **Analyse → Compare Means → Paired-Samples T-Test**
 - Cung cấp cặp biến phối hợp
 - Điều chỉnh độ tin cậy (nếu cần)...

V.D., TRUNG BÌNH CỦA 2 TỔNG THỂ - TỪNG CẶP



TRUNG BÌNH CỦA 2 TỔNG THỂ - TỪNG CẶP (SPV)

Giả thuyết H_0 : Cảm nhận của khách hàng là không đổi

Paired Samples Correlations				
	N	Correlation	Sig.	
Pair 1 beforeSLA & afterSLA	50	.398	.004	

Paired Samples Test				
	Pair 1			
	beforeSLA - afterSLA			
Mean	-.460			
Std. Deviation	1.034			
Std. Error Mean	.146			
95% Confidence Interval of the Difference	Lower Upper			
	-1.754 -1.166			
t	-3.145			
df	49			
Sig. (2-tailed)	.003			

Mức ý nghĩa = 0.00 < 0.05

Bác bỏ giả thuyết H_0

KẾT LUẬN: Cảm nhận của khách hàng có thay đổi

IBM®

SPSS®

PHÂN TÍCH TƯƠNG QUAN

Tương quan 2 biến (Bivariate)

1. Tương quan Pearson
2. Tương quan Spearman

1. TƯƠNG QUAN 2 BIẾN

■ Tương quan Pearson

- Áp dụng trong trường hợp mẫu có phân phối chuẩn
- Lượng hóa mối liên hệ tuyến tính giữa 2 biến định lượng
- Mối liên hệ giữa 2 biến có thể **$[-1, 1]$**
 - $r > 0$: liên hệ tuyến tính thuận
 - $r < 0$: liên hệ tuyến tính nghịch
 - $r = 0$: không có liên hệ hoặc liên hệ phi tuyến
- Mối liên hệ này có tính chất đối xứng.

r

TƯƠNG QUAN 2 BIẾN

- Hệ số tương quan giữa hai biến x và y:

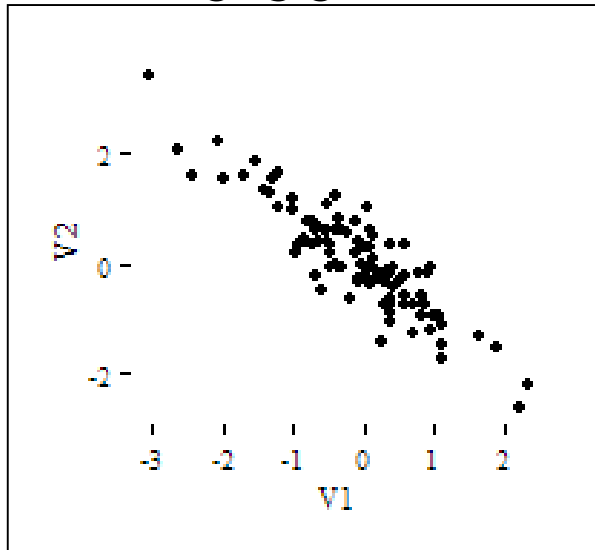
$$r_{xy} = \frac{cov(x, y)}{(\sqrt{var(x)}) * (\sqrt{var(y)})}$$

Trong đó:

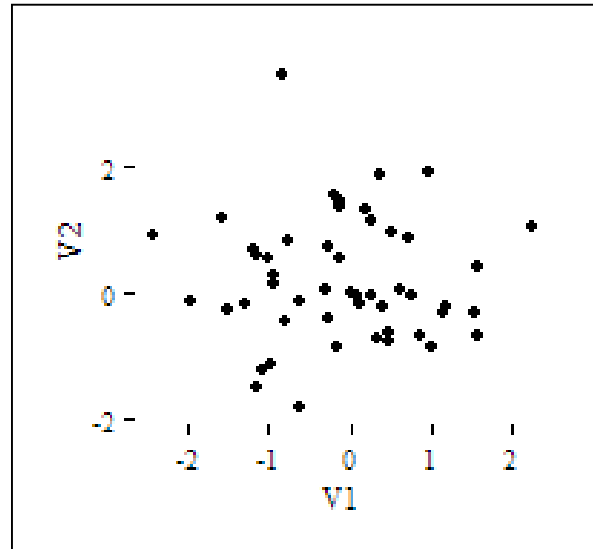
- $cov(x, y)$: hiệp phương sai của x và y
- $var(x)$: phương sai của x
- $var(y)$: phương sai của y

TƯƠNG QUAN 2 BIẾN

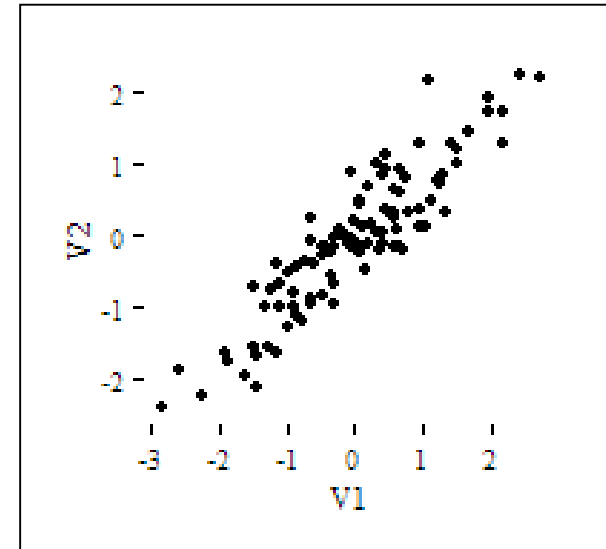
$$r = -0.90$$
$$= 0.90$$



$$r = 0.00$$



$$r$$



A. TƯƠNG QUAN PEARSON

- Giả thuyết: H_0 và H_1 , và hệ số tương quan (r)
 - Kiểm định 2 phía (two-tailed)
 - $H_0: r = 0$ - không có mối tương quan
 - $H_1: r \neq 0$ - có mối tương quan
 - Kiểm định 1 phía (one-tailed)
 - $H_0: r = 0$ - không có mối tương quan
 - $H_1: r > 0$ - có mối tương quan thuận
 - $H_1: r < 0$ - có mối tương quan nghịch

TƯƠNG QUAN PEARSON

- Một số yêu cầu về dữ liệu:
 - 2 biến liên tục (thang đo khoảng/tỷ lệ)
 - Các quan sát có giá trị trên cả 2 biến
 - Các quan sát độc lập
 - Các biến có phân phối chuẩn
 - Không có trường hợp bất thường...

TƯƠNG QUAN PEARSON

■ Ví dụ,

- *Mục tiêu nghiên cứu: Có mối quan hệ như thế nào giữa độ tuổi và thu nhập.*
- *Câu hỏi nghiên cứu: Độ tuổi và thu nhập có mối quan hệ với nhau hay không?*
- *Các biến: Độ tuổi và thu nhập (2 biến định lượng).*
- *H_0 : “Độ tuổi và thu nhập không có liên hệ với nhau”*

TƯƠNG QUAN PEARSON

- Quy trình thực hiện
 - **Analyze → Correlate → Bivariate**
 - Chọn biến vào **Variables**
 - **Tại Correlation Coefficients → Pearson**
→ Test of Significance → OK...

V.D., TƯƠNG QUAN PEARSON

The image shows the SPSS Data Editor window with a dataset named 'data_Correlation.SAV'. The 'Data View' tab is active, displaying a table with two columns: 'Input_Score' and 'Output_Score'. The first row of data is highlighted. A 'Bivariate Correlations' dialog box is open, showing the variables 'Input_Score' and 'Output_Score' in the 'Variables:' list. Under the 'Correlation Coefficients' section, the 'Pearson' checkbox is checked and highlighted with a red rectangle. The 'Test of Significance' section shows 'Two-tailed' selected. The 'Flag significant correlations' checkbox is also checked. The 'OK' button is visible at the bottom of the dialog box.

	Input_Score	Output_Score
1	28.00	7.40
2	28.50	7.78
3	21.25	5.21
4	22.75	6.96
5	29.50	6.89
6	28.75	6.64
7	21.75	7.10
8	22.00	7.50
9	20.50	4.70
10	20.00	6.00
11	28.00	9.22
12	22.75	5.68
13	21.50	6.04
14	20.75	6.20

V.D., TƯƠNG QUAN PEARSON (SPV)

H_0 : Độ tuổi và thu nhập không có liên hệ với nhau

→ **Correlations**

Hệ số tương quan giữa
2 biến là $r = 0.670$

Mức ý nghĩa $sig. = 0.000 < 0.05$

Bác bỏ giả thuyết H_0

Correlations

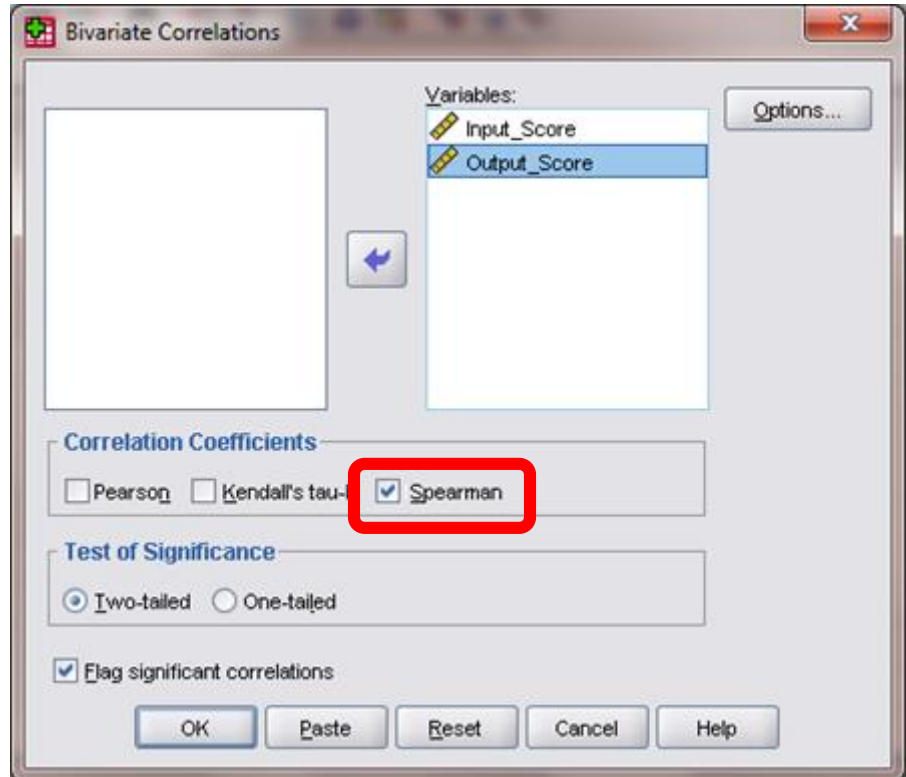
	Input_Score	Output_Score
Input_Score	1	.670**
	300	300
Output_Score	.670**	1
	300	300

** . Correlation is significant at the 0.01 level (2-tailed).

KẾT LUẬN: Độ tuổi và thu nhập có liên hệ với nhau

B. TƯƠNG QUAN SPEARMAN

- Tương quan - Spearman
 - Rank Correlation Coefficient
 - Áp dụng trong trường hợp tổng thể không có phân phối chuẩn...



IBM®

SPSS®

PHÂN TÍCH HỒI QUY

1. Hồi quy tuyến tính đơn biến
2. Hồi quy tuyến tính đa biến (bội)

$$Y = \beta_0 + \sum (\beta_i * X_i)$$

PHÂN TÍCH HỒI QUY

- Mô hình hóa mối quan hệ tuyến tính giữa 1 biến phụ thuộc (Y) với nhiều biến độc lập (X_i)
- Không có tính đối xứng như tương quan
- Ví dụ,
Các yếu tố ảnh hưởng đến chất lượng KTX BUIH...

1. HỒI QUY TUYẾN TÍNH ĐƠN

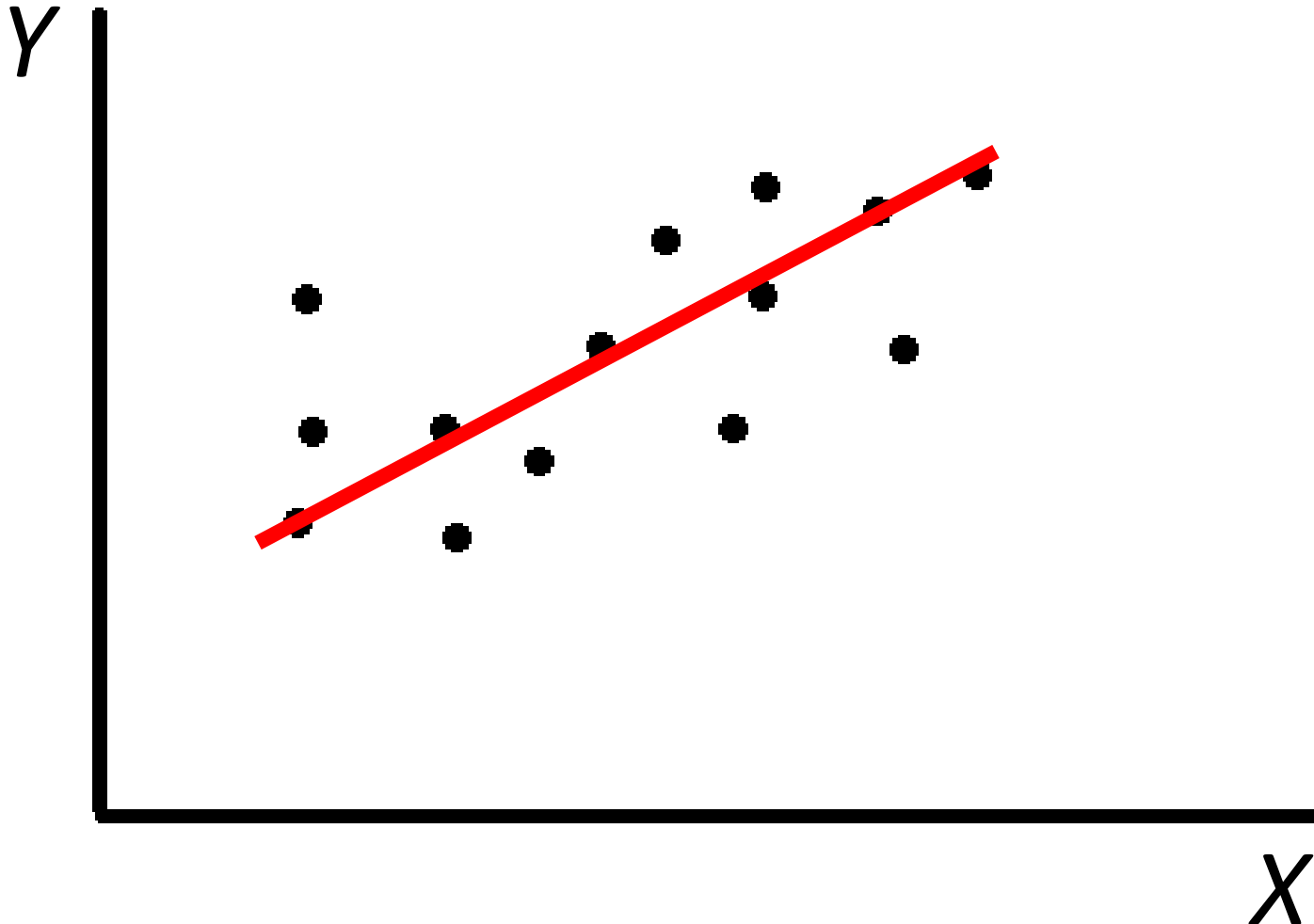
- Mô hình tổng quát

$$Y_i = \beta_0 + \beta_1 * X_i + \varepsilon_i$$

Trong đó

- Y_i : giá trị dự đoán thứ i của biến phụ thuộc
- X_i : giá trị quan sát thứ i của biến độc lập
- β_0 và β_1 : các hệ số hồi quy - theo phương pháp OLS
- ε_i : phần dư, biến độc lập ngẫu nhiên $N(0, \sigma^2)$...

V.D: ĐỒ THỊ HỒI QUY TUYẾN TÍNH ĐƠN



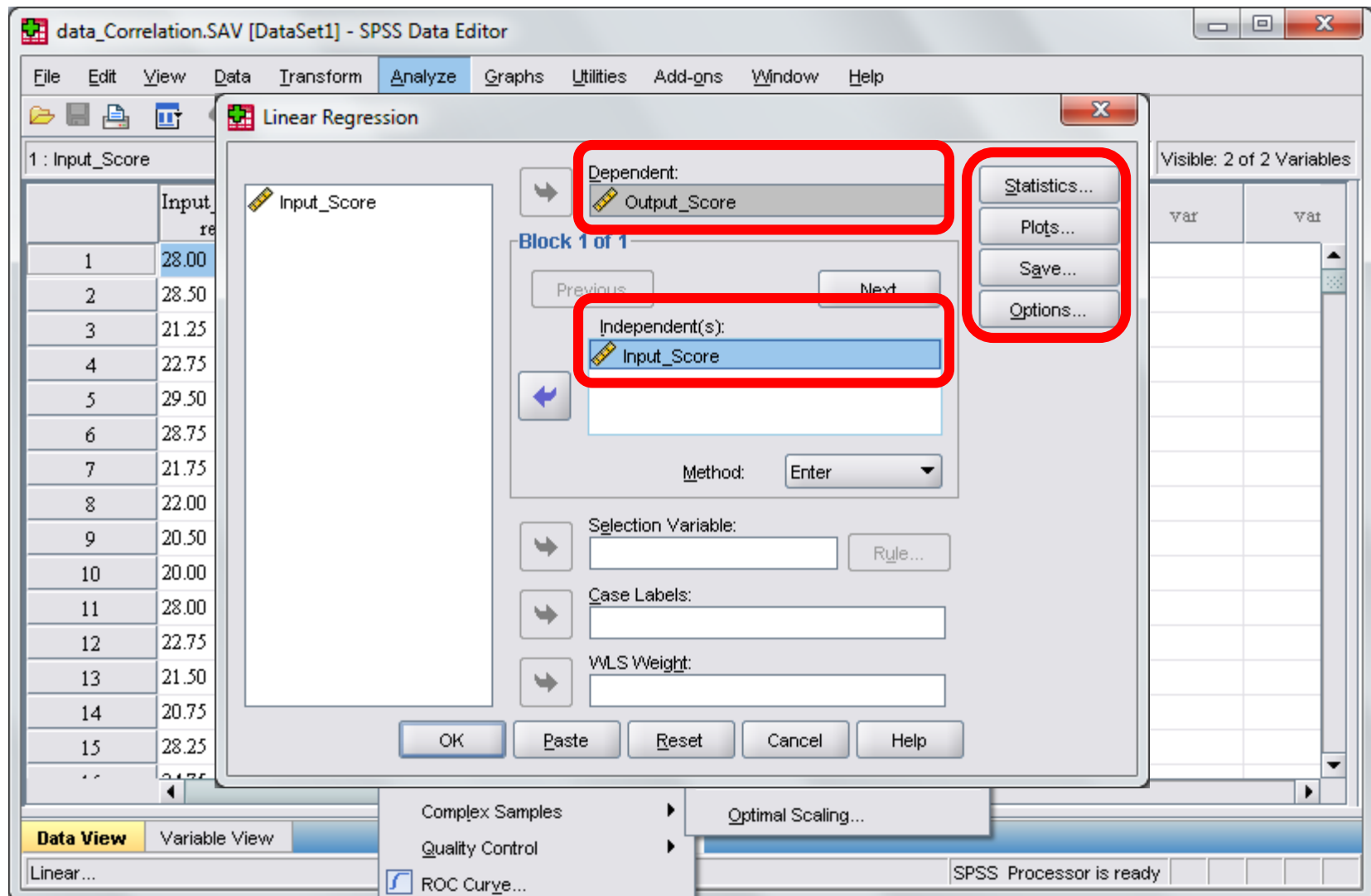
HỒI QUY TUYẾN TÍNH ĐƠN

- Một số giả định và yêu cầu về dữ liệu
 - 2 biến liên tục (thang đo khoảng/tỷ lệ)
 - Các quan sát có giá trị trên cả 2 biến
 - Tồn tại quan hệ tuyến tính giữa 2 biến
 - Các giá trị của Y độc lập với nhau
 - Phân phối chuẩn của Y với phương sai không đổi
 - Các giá trị trung bình $\mu(Y|X)$ nằm trên đường thẳng
 - Phần dư có phân phối chuẩn...

HỒI QUY TUYẾN TÍNH ĐƠN

- Quy trình thực hiện
 - **Analyze** → **Regression** → **Linear...**
 - Chọn biến vào **Dependent** và **Independent**
 - Kiểm định các giả định → **Statistics**
 - Biểu đồ → **Plots...**
 - **OK...**

V.D: HỒI QUY TUYẾN TÍNH ĐƠN



V.D., HỒI QUY TUYẾN TÍNH ĐƠN (SPV)

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Input_Score ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: Output_Score

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.670 ^a	.449	.448	.70223

a. Predictors: (Constant), Input_Score

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	119.986	1	119.986	243.320	.000 ^a
	Residual	146.950	298	.493		
	Total	266.936	299			

Hệ số hồi quy

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.974	.349		2.788	.006
	Input_Score	.232	.015	.670	15.599	.000

a. Dependent Variable: Output_Score

Phương pháp nhập một lần

Đo lường mức độ phù hợp của mô hình đối với mẫu (R^2)

Mức độ phù hợp của mô hình với tổng thể (F lớn, $sig. < 0.05$)

Mức ý nghĩa của các hệ số hồi quy $sig. = 0.00 < 0.05$

$$Y_i = 0.974 + 0.232 * X_i + \varepsilon$$

HỒI QUY TUYẾN TÍNH ĐƠN (SPV)

■ Variables Entered/Removed

- **Model:** cho phép chạy nhiều mô hình mỗi lần thực hiện hồi quy, cột Model là số thứ tự của mô hình.
- **Variables Entered:** cho phép nhập các biến vào theo khối (blocks) và thực hiện hồi quy theo từng bước (stepwise)
- **Variables Removed:** liệt kê các biến bị loại bỏ khỏi hồi quy, cột này thường trống trừ khi thực hiện hồi quy theo từng bước.
- **Method:** phương pháp. V.d., từng bước

Variables Entered/Removed ^b			
Model	Variables Entered	Variables Removed	Method
1	Input_Score ^a	.	Enter
a. All requested variables entered.			
b. Dependent Variable: Output_Score			

HỒI QUY TUYẾN TÍNH ĐƠN (SPV)

■ Model Summary

- **Model:** số thứ tự của mô hình hồi quy
- **R:** tương quan giữa biến độc lập và biến phụ thuộc
- **R-Squared:** mức độ biến thiên của biến phụ thuộc được giải thích bởi biến độc lập
- **Adjusted R-Square:** mức độ biến thiên của biến phụ thuộc được giải thích bởi biến độc lập
 - V.d., 44.8% ($R^2 = 0.448$) sự biến thiên của chất lượng KTX được giải thích bởi chất lượng nhân viên KTX.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.670 ^a	.449	.448	.70223
a. Predictors: (Constant), Input_Score				

R^2 VÀ R^2 HIỆU CHỈNH

- **R-Squared** tăng khi thêm 1 biến độc lập (X) vào mô hình, dù biến thêm vào không có ý nghĩa thống kê
- **Adjusted R-Squared** chỉ tăng khi biến độc lập (X) có ý nghĩa thống kê và ảnh hưởng đến biến phụ thuộc (Y)
- **R-Squared** không có giá trị âm
- **Adjusted R-Squared** có thể âm khi R-Squared gần giá trị 0...

R^2

HỒI QUY TUYẾN TÍNH ĐƠN (SPV)

■ ANOVA

- **Sum of squares:** tổng bình phương
- **df:** bậc tự do, Residual = $n - 1$...
- **Mean Square:** Sum of Squares/df
- **F và Sig.:** kiểm định F và p-value...
 - $F = \text{Mean Square (Regression)} / \text{Mean Square (Residual)}$.
 - Giả thuyết H_0 : “tất cả các hệ số hồi quy của mô hình bằng 0”
 - Sig. $< \alpha$: mô hình hồi quy có ý nghĩa thống kê.

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	119.986	1	119.986	243.320	.000 ^a
	Residual	146.950	298	.493		
	Total	266.936	299			

a. Predictors: (Constant), Input_Score
b. Dependent Variable: Output_Score

HỒI QUY TUYẾN TÍNH ĐƠN (SPV)

■ Coefficients

- **B**: các hệ số hồi quy tuyến tính.
- **Std. Error**: sai số chuẩn cho các hệ số
- **Beta**: các hệ số được chuẩn hóa.
- **t** và **Sig.**: Kiểm định t và p-value
 - Sig. > α : không có ý nghĩa thống kê (loại bỏ b/beta tương ứng)
 - Sig. < α : có ý nghĩa thống kê (giữ lại b/beta tương ứng)...

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	
1	(Constant)	.974	.349		2.788 .006
	Input_Score	.232	.015	.670	15.599 .000

a. Dependent Variable: Output Score

IBM®

SPSS®

2. HỒI QUY TUYẾN TÍNH BỘI

- Mô hình tổng quát

$$Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \dots + \beta_p * X_{pi} + \varepsilon_i$$

Trong đó

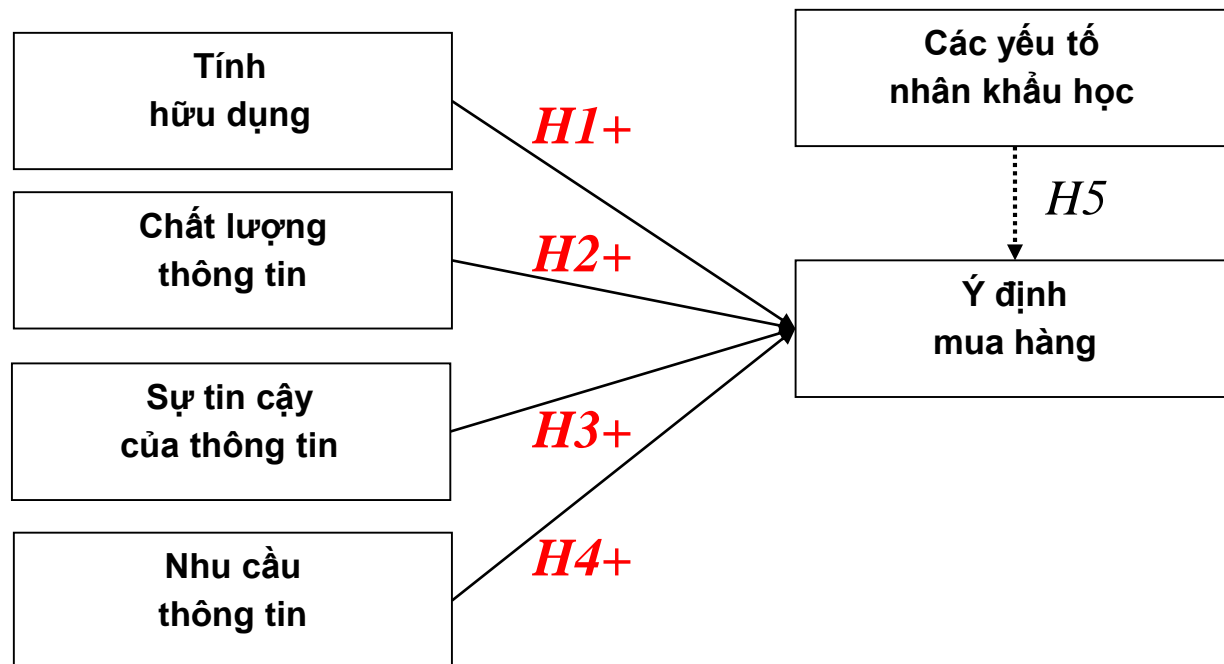
- Y_i : giá trị dự đoán thứ i của biến phụ thuộc
- X_{ki} : giá trị quan sát thứ i của biến độc lập thứ k
- β_i : các hệ số hồi quy riêng phần - theo phương pháp OLS
- ε_i : phần dư, biến độc lập ngẫu nhiên $N(0, \sigma^2)$.

HỒI QUY TUYẾN TÍNH BỘI

- Một số giả định và yêu cầu về dữ liệu
 - Các biến độc lập - phụ thuộc liên tục (thang đo khoảng/tỷ lệ)
 - Các quan sát có giá trị độc lập
 - Tồn tại quan hệ tuyến tính giữa các biến độc lập - phụ thuộc
 - Các giá trị của Y độc lập với nhau
 - Phân phối chuẩn của Y với phương sai không đổi
 - Không tồn tại đa cộng tuyến giữa các biến độc lập
 - Phần dư có phân phối chuẩn...

V.D: HỒI QUY TUYẾN TÍNH BỘI

- Xác định mức độ tác động của các yếu tố có ảnh hưởng đến ý định mua nước hoa ở TP. HCM.*



HỒI QUY TUYẾN TÍNH BỘI

■ Quy trình thực hiện

- A. Xem xét ma trận hệ số tương quan
- B. Xây dựng phương trình hồi quy
- C. Kiểm định các giả thuyết
- D. Giải thích các hệ số hồi quy
- E. Xác định tầm quan trọng của các biến
- F. Lựa chọn các biến độc lập cho mô hình
- G. Dò tìm các vi phạm giả định cần thiết [5]...

A. MA TRẬN HỆ SỐ TƯƠNG QUAN

The screenshot displays the SPSS Data Editor window for a file named *data_Correlation.SAV [DataSet1]. The main window shows a data table with 16 rows and 3 columns: SelfControl, Input_Score, and Output_Score. The 'Bivariate Correlations' dialog box is open, showing the variables selected for analysis: Input_Score, SelfControl, and Output_Score. The dialog box also shows options for correlation coefficients (Pearson, Kendall's tau-b, Spearman) and tests of significance (Two-tailed, One-tailed). The 'Flag significant correlations' checkbox is checked.

	SelfControl	Input_Score	Output_Score
1	28.00	28.00	
2	28.50	28.50	
3	21.25	21.25	
4	22.75	22.75	
5	29.50	29.50	
6	28.75	28.75	
7	21.75	21.75	
8	22.00	22.00	
9	20.50	20.50	
10	20.00	20.00	
11	28.00	28.00	
12	22.75	22.75	
13	21.50	21.50	
14	20.75	20.75	
15	28.25	28.25	
16	24.75	24.75	

Bivariate Correlations

Variables:

- Input_Score
- SelfControl
- Output_Score

Correlation Coefficients

☒ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance

☒ Two-tailed ☐ One-tailed

☒ Flag significant correlations

OK Paste Reset Cancel Help

MA TRẬN HỆ SỐ TƯƠNG QUAN (SPV)

→ Correlations

Hệ số tương quan giữa hai biến độc lập thấp & không có ý nghĩa

[DataSet1] D:\02 Tai lieu dao tao\002 Tin hoc ung dung\data_Corre

Hệ số tương quan giữa biến phụ thuộc và các biến độc lập tương đối cao & có ý nghĩa

		Correlations		
		Input_Score	SelfControl	Output_Score
Input_Score	Pearson Correlation	1	.070	.374**
	Sig. (2-tailed)		.229	.000
	N	300	300	300
SelfControl	Pearson Correlation	.070	1	.784**
	Sig. (2-tailed)	.229		.000
	N	300	300	300
Output_Score	Pearson Correlation	.374**	.784**	1
	Sig. (2-tailed)	.000	.000	
	N	300	300	300

** . Correlation is significant at the 0.01 level (2-tailed).

KẾT LUẬN: Có thể sử dụng các biến độc lập trong mô hình

B. XÂY DỰNG MÔ HÌNH HỒI QUY

- Quy trình thực hiện
 - **Analyze** → **Regression** → **Linear...**
 - Chọn biến vào **Dependent** và các biến **Independent**
 - Kiểm định các giả định → **Statistics**
 - Biểu đồ → **Plots...**
 - **OK...**

XÂY DỰNG MÔ HÌNH HỒI QUY

data_Correlation.SAV [DataSet1] - SPSS Data Editor

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

Linear Regression

1 : Input_Score

	Input_Score
1	28.00
2	28.50
3	21.25
4	22.75
5	29.50
6	28.75
7	21.75
8	22.00
9	20.50
10	20.00
11	28.00
12	22.75
13	21.50
14	20.75
15	28.25
16	24.75

Linear Regression: Statistics

Dependent: Input_Score

Independent: SelfControl

Regression Coefficient

- ☒ Estimates
- ☐ Confidence intervals
- ☐ Covariance matrix

Residuals

- ☐ Durbin-Watson
- ☐ Casewise diagnostics
- ☒ Outliers outside: 3 standard deviations
- ☐ All cases

Model fit

- ☐ R squared change
- ☐ Descriptives
- ☐ Part and partial correlations
- ☐ Collinearity diagnostics

Continue Cancel Help

Complex Samples

Quality Control

ROC Curve...

Optimal Scaling...

SPSS Processor is ready

PHƯƠNG TRÌNH HỒI QUY

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.847 ^a	.717	.715	.52385

a. Predictors: (Constant), SelfControl, Input_Score

Hệ số hồi quy

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	2.323	.265		8.762	.000			
	Input_Score	.115	.011	.321	10.361	.000	.374	.515	.320
	SelfControl	.677	.028	.762	24.613	.000	.784	.819	.760

a. Dependent Variable: Output_Score

Hệ số xác định tổng thể $R^2 = 0.717$, nên các biến độc lập giải thích được khoảng 71,7 % sự biến thiên của biến phụ thuộc

$$Y = 2.232 + 0.115 * X_1 + 0.677 * X_2 + \varepsilon$$

C. KIỂM ĐỊNH CÁC GIẢ THUYẾT (SPV)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	206.456	2	103.228	376.172	.000 ^a
	Residual	81.502	297	.274		
	Total	287.958	299			

a. Predictors: (Constant), SelfControl, Input_Score

b. Dependent Variable: Output_Score

Mức độ phù hợp của mô hình với tổng thể (F lớn, *sig.* < 0.05)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	2.323	.265		8.762	.000			
	Input_Score	.115	.011	.321	10.361	.000	.374	.515	.320
	SelfControl	.677	.028	.762	24.613	.000	.784	.819	.760

a. Dependent Variable: Output_Score

Các biến đều có ý nghĩa trong mô hình (*sig.* < 0.05)

KẾT LUẬN: Các giả thuyết đều được ủng hộ (không bác bỏ)

E. TẦM QUAN TRỌNG CÁC BIẾN (SPV)

The image shows the SPSS Data Editor window with a dataset named 'data_Correlation.SAV'. The 'Analyze' menu is open, and the 'Linear Regression' dialog box is displayed. The 'Statistics' sub-dialog box is active, showing the 'Regression Coefficient' section. The 'Part and partial correlations' checkbox is checked and highlighted with a red rectangle. A red callout box with white text points to this checkbox, stating: 'Xác định tầm quan trọng của các biến trong mô hình' (Determine the importance of variables in the model). The 'Residuals' section is also visible, with 'Outliers outside: 3 standard deviations' selected. The 'Data View' tab is active at the bottom, showing a table with 16 rows and 2 columns: 'Input_Score' and 'SelfControl'.

	Input_Score
1	28.00
2	28.50
3	21.25
4	22.75
5	29.50
6	28.75
7	21.75
8	22.00
9	20.50
10	20.00
11	28.00
12	22.75
13	21.50
14	20.75
15	28.25
16	24.75

TẦM QUAN TRỌNG CÁC BIẾN (SPV)

2 biến này có ảnh hưởng đối với mô hình lớn hơn các biến còn lại

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
	B	Std. Error	Beta			Zero-order	Partial	Part
1 (Constant)	-.690	.282		-2.449	.014			
AGE	.008	.007	.021	1.133	.257	.062	.020	.020
GENDER	.642	.119	.094	5.388	.000	.103	.096	.093
MARITAL	-.502	.121	-.075	-4.147	.000	-.047	-.074	-.072
INCOME	-.071	.021	-.161	-3.412	.001	.149	-.061	-.059
LIMIT	.055	.008	.327	6.879	.000	.183	.122	.119
OCCUP	.334	.057	.103	5.889	.000	.121	.105	.102
GUARTY	.321	.118	.048	2.732	.006	.056	.049	.047

a. Dependent Variable: OVERDUE

F. LỰA CHỌN CÁC BIẾN ĐỘC LẬP

- Mục đích

- Lựa chọn hiệu quả các biến đưa vào mô hình

- Thực hiện

- Tăng dần số lượng các biến và kiểm tra mức độ phù hợp của mô hình...

LỰA CHỌN CÁC BIẾN ĐỘC LẬP

data_Correlation.SAV [DataSet1] - SPSS Data Editor

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

1 : Input_Score

	Input_Score
1	28.00
2	28.50
3	21.25
4	22.75
5	29.50
6	28.75
7	21.75
8	22.00
9	20.50
10	20.00
11	28.00
12	22.75
13	21.50
14	20.75
15	28.25
16	24.75

Linear Regression

Linear Regression: Statistics

Regression Coefficient

- ☒ Estimates
- ☐ Confidence intervals
- ☐ Covariance matrix
- ☒ Model fit
- ☒ R squared change
- ☐ Descriptives
- ☐ Part and partial correlations
- ☐ Collinearity diagnostics

Residuals

- ☐ Durbin-Watson
- ☐ Casewise diagnostics
- ☒ Outliers outside: 3 standard deviations
- ☐ All cases

Continue Cancel Help

OK Paste Reset Cancel Help

Plots... Save... Options...

Data View Variable View

SPSS Processor is ready

Lựa chọn các biến độc lập

LỰA CHỌN CÁC BIẾN ĐỘC LẬP (SPV)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Hệ số xác định tổng thể (R^2) thay đổi				
					R Square Change	F Change	df1	df2	Sig.
1	.374 ^a	.140	.137	.91179	.140	48.367	1	298	.000
2	.847 ^b	.717	.715	.52385	.577	605.813	1	297	.000

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40.211	1	40.211	48.367	.000 ^a
	Residual	247.747	298	.831		
	Total	287.958	299			
2	Regression	206.456	2	103.228	376.172	.000 ^b
	Residual	81.502	297	.274		
	Total	287.958	299			

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.507	.454		7.730	.000
	Input_Score	.134	.019	.374	6.955	.000
2	(Constant)	2.323	.265		8.762	.000
	Input_Score	.115	.011	.321	10.361	.000
	SelfControl	.677	.028	.762	24.613	.000

F2 >> F1

KẾT LUẬN: Mô hình 2 biến tốt hơn hẳn so với mô hình 1 biến

G. HIỆN TƯỢNG ĐA CỘNG TUYẾN

Xem xét hiện tượng đa cộng tuyến (VIF)

The screenshot shows the SPSS Data Editor window with the 'Statistics' dialog box open. The 'Collinearity diagnostics' checkbox is selected and highlighted with a red box. The 'Residuals' section also shows 'Durbin-Watson' selected and 'Casewise diagnostics' with 'Outliers outside: 3 standard deviations' selected. The 'Data View' tab is active at the bottom.

Variable	Value
1	28.50
2	28.50
3	21.25
4	22.75
5	29.50
6	28.75
7	21.75
8	22.00
9	20.50
10	20.00
11	28.00
12	22.75
13	21.50
14	20.75
15	28.25
16	24.75

SPSS Processor is ready

HIỆN TƯỢNG ĐA CỘNG TUYẾN (SPV)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	206.456	2	103.228	376.172	.000 ^a
	Residual	81.502	297	.274		
	Total	287.958	299			

a. Predictors: (Constant)

b. Dependent Variable: Output_Score

Nếu $VIF < 10$: Các biến độc lập không có tương quan với nhau

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.323	.265		8.762	.000		
	Input_Score	.115	.011	.321	10.361	.000	.995	1.005
	SelfControl	.677	.028	.762	24.613	.000	.995	1.005

a. Dependent Variable: Output_Score

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Input_Score	SelfControl
1	1	2.873	1.000	.00	.00	.02
	2	.120	4.891	.02	.02	.98
	3	.007	20.662	.98	.98	.00

a. Dependent Variable: Output_Score

KẾT LUẬN: Không xảy ra hiện tượng đa cộng tuyến

IBM®

SPSS®

Q&A

HẾT