

# XỬ LÝ SỐ LIỆU ĐIỀU TRA DÂN SỐ

## bằng chương trình SPSS 10.0

Phan Đắc Lộc  
Vụ Dân số - Lao động

### I. SPSS: Phần mềm xử lý các cuộc điều tra mẫu thống kê

#### 1.1. Ưu điểm:

- Xử lý nhanh/ gọn các cuộc điều tra mẫu thống kê
- Chuyển dữ liệu từ EXCEL, LOTUS, ISSA... sang SPSS và ngược lại rất dễ dàng
- Dùng tổng hợp, phân tích thống kê chuyên sâu.

#### 1.2. Nhược điểm: Hạn chế bản ghi:

- Số biến được lập trong Data file không quá 300 biến; có nghĩa là nếu bảng hỏi có quá nhiều câu hỏi, trên 300 câu thì không đủ để lập hết được.
- Một dòng dùng để nhập tin cho 1 case (1 bản ghi: người, nhà..) không quá 10 triệu, có nghĩa là với cuộc điều tra mẫu lớn hơn 10 triệu dân thì không đủ chỗ lập hết được trong 1 Data file mà phải lập tách ra sau đó ghép lại.

**Ví dụ:** Trong Data file có tên “ Người” từ kết quả 3 % mẫu của TCTK, số dòng tương ứng với số dân điều tra mẫu là: 2368167 (cases - bản ghi người). Số cột tương ứng với số câu hỏi được mã hoá trên phiếu điều tra là: 47 (Variables – biến).

**Cấu trúc các file trong SPSS:** Trong xử lý các bài toán Thống kê, có 3 dạng file thường gặp:

a/ File dữ liệu (Data file): file này có đuôi (\*. **SAV**). File dữ liệu có thể được tạo lập mới từ đầu, có thể sửa lại được và có thể chuyển từ file đã lập trên EXCEL, LOTUS, ISSA.. sang SPSS

b/ file các lệnh xử lý (Systat): File có đuôi (\*. **SPS**). Phần lệnh lập trình được lưu lại trong SPSS, Từ đây có thể kiểm tra các lệnh đã làm đúng hay sai và sửa lại.

c/ file lưu kết quả xử lý/ tổng hợp (Output): File có đuôi (\*. **SPO**)

Từ kết quả tổng hợp bằng SPSS, chúng ta có thể copy file sang các chương trình khác như: Word, Excel có đuôi (\*.xls); LOTUS có đuôi (\*. W); Dbase có đuôi (\*.dbf)... để chỉnh sửa, vẽ đồ thị, in ấn kết quả.

Dưới đây chỉ trình bày chi tiết 3 dạng chính của cấu trúc file SPSS liên quan đến việc khai thác kết quả điều tra mẫu 3 % trong TĐTDS và nhà ở 1999 do TTTTCTW đã lập.

### II. Tạo biến trong Data file

#### 1. Tạo biến mới:

Vào:  =>  =>

màn hình hiển thị một Data file mới, để tạo các biến mới phải vào cửa sổ dưới có chữ:

Sau đó chúng ta tạo lập biến mới lần lượt theo các bước:

Đặt tên biến (name): Dùng bàn phím đánh trực tiếp tên biến cần đặt, ví dụ: TĐHV,GT, DT, TONGIAO,...

- Kiểu biến (type): có 8 dạng biến: Dạng số, dạng chữ, dạng tiền tệ Dolar, ngày/tháng...Thực tế thường dùng 2 dạng sau:

+ Chuỗi ký tự: Phải đặt số ký tự - (Charasters - )

**Ví dụ:** Muốn nhập biến “Họ và tên” thì phải dành khoảng 20 ký tự

+ Dạng số - mã số (numeric): Phải đặt độ rộng số (Width - ) và phải đặt số lẻ thập phân (Decimal - )

Ví dụ: + Biến “Ngành đào tạo”- có mã 6 số thì phải đặt độ rộng biến ít nhất là , số lẻ thập phân là

+ Biến “Ngành Dân tộc”- có mã 2 số thì phải đặt độ rộng biến ít nhất là , số lẻ thập phân là

- Làm nhãn tên biến (Label):

Ví dụ: TĐHV - nhãn là: “Trình độ học vấn”, GT- nhãn là: “Giới tính”

DT - nhãn là: “Dân tộc”; P16 - nhãn là: “Hoạt động kinh tế 12 tháng”...

- Tên nhãn biến và các giá trị của biến (Value): đây chính là mã số của các câu trả lời cho 1 câu hỏi ghi trên phiếu điều tra.

Ví dụ: Biến GT - “Giới tính” nhận các giá trị: 1= nam, 2 = nữ ;

Biến DT - “Dân tộc” nhận các giá trị : 01 = Kinh, 02 = Tày....

2. Các biến đã lập trong bản ghi (Data file) kết quả mẫu 3% TĐTDS và nhà ở 1999:

Vào:  =>  =>

Mở từng bản ghi và sau đó vào:

2.1 Bản ghi “Người”: có 47 biến - xem ở phần

Bản ghi “Nhà”: có 21 biến

Bản ghi “Chết”: có 12 biến

Tóm lại:

+ Các biến tương ứng với số câu hỏi có trên phiếu điều tra

+ Các giá trị biến tương ứng với mã số các câu trả lời trên phiếu điều tra.

Ví dụ cụ thể:

Biến P1:

Name (tên): Province (tỉnh)

Type (kiểu): numeric (chữ số)

Width (số ký tự): 3 chữ số

Decimals (phần thập phân): 0

Label (nhãn): mã tỉnh/thành phố

Values (giá trị): Nhấn chuột vào phần  trong ô:

101 = “Hà Nội”

103 = “Hải Phòng”

105 = “Hà Tây”

....

cuối cùng là 823 = “Cà Mau”

Missing: (giá trị sót): 0

Column (cột): độ rộng cột biến = 8

Align (đặt): - Right - dồn bên phải;

- Left - trái; Center - giữa

Biến P16:

Name (tên): P16

Type (kiểu): numeric (chữ số)

Width (số ký tự): 1 chữ số

Decimals (phần thập phân): 0

Label (nhãn): Main activity 12 months

- Hoạt động chính trong 12 tháng

Values (giá trị): Nhấn chuột vào phần  trong ô:

1 = “Working – Làm việc”

2 = “HouseWork – Nội trợ”

3 = “Studying - đang đi học”

4 = “Invalid - Mất khả năng lao động”

5 = “Unemployment - Thất nghiệp”

6 = “Without demand of job - Không có nhu cầu làm việc”

7 = “NR - không xác định câu trả lời”

3. Chèn thêm biến và thêm case vào file đã lập sẵn:

Muốn tạo chỗ cho 1 biến (cột) mới:

Vào  =>

Muốn tạo chỗ cho 1 case (dòng) mới:

Vào  =>

4. Tạo thêm biến mới từ biến đã lập:

Thực tế khi phân tích số liệu, nhiều tác giả phải tạo ra các biến mới từ các

biến đã có sẵn trong Data file; chẳng hạn: nhóm lại tuổi, nhóm lại trình độ học vấn theo cấp học, nhóm các ngành/ nghề theo một số nhóm chính.v.v..

Ví dụ: Từ biến “Age” nhóm thành “Age groups”

Từ biến “P111” - lớp học nhóm thành “TĐHV theo cấp”...

Cách làm: Vào **Transform=> Recode**  
=> **into same Variable/into defferent Variable**

Chọn: **into same Variable**

Ví dụ: Nhóm lại biến “Age” thành “Age groups” thì chọn biến “Age”. Khi đó màn hình hiện:

**Old and New Variable**

Old Value	New Value
(*) Value: 0	(*) Value: 0
	←→ vẫn giữ 0 tuổi

(\*) Range

(Nhóm tuổi)	New Value	Nhãn
1 through 4	→ 1	“1 - 4 “ tuổi
5 through 9	→ 2	“5 - 9 “ tuổi
10 through 14	→ 3	“10 - 14 “ tuổi
15 through 19	→ 4	“15 - 19 “ tuổi

Tiếp tục cho đến nhóm cuối cùng 75 + (hoặc 80+) ... thì cho vào phần (\*) Range **[.....] through highest.**

Sau khi chạy, SPSS sẽ tạo lập cho ta 1 biến mới là “Age groups” gồm các nhóm: 0 tuổi, 1-4 tuổi, 5-9 tuổi, 10-14 tuổi, ..... ; cuối cùng là 75+.

Việc nhóm lại biến như thế nào là do các tác giả tự đặt ra theo mục đích nghiên cứu.

### III. Lọc các Cases

+ Bình thường, khi làm các lệnh tổng hợp, nếu không lọc các **Case** thì đầu ra sẽ cho ta kết quả tổng hợp của toàn bộ các Case được lập trong data file.

+ Trong thực tế, theo mục đích nghiên cứu/ phân tích, có biểu tổng hợp chỉ liên quan đến số phụ nữ 15-49 tuổi, hoặc số lao động 13 tuổi trở lên, hoặc hôn nhân chỉ nghiên cứu đối với người từ 18 tuổi trở lên .v.v.. Do vậy, SPSS cho phép ta lọc các Case theo các biến phù hợp với các đối tượng nêu trên trước khi tổng hợp.

Thao tác vào:

**Data=>Select cases=>If condition is satifed**  
=> **If..**

Khi đó mà hình hiện ra bảng chỉ thị đặt điều kiện lọc, chuyển các biến cần lọc và điều kiện lọc từ cửa sổ trái màn hình sang góc trên bên phải màn hình:

Các biến trong Data file	=> chuyển các biến cần lọc và đặt điều kiện	
	Bản phím dùng để đặt các điều kiện lọc	Các hàm số liên quan đến điều kiện lọc

Sau khi đặt xong điều kiện, bấm **Continue (tiếp tục)**

Tiếp theo phía dưới: **Unselected cases are:**

+ Nếu bấm **Filtered** thì kết quả sẽ lọc gạch chéo các Case không thỏa mãn điều kiện lọc;

+ Nếu bấm **Deleted** thì kết quả sẽ loại các Case không thỏa mãn điều kiện lọc ra khỏi Data file, chỉ giữ lại các Case thỏa mãn điều kiện lọc.

Sau đó bấm: **Paste** , Các lệnh điều kiện lọc sẽ được lưu vào file Syntat (\*. sys)

Cuối cùng, bấm **Run** , máy bắt đầu thực hiện các lệnh lọc.

Nếu không bấm **Paste** mà bấm ngay **OK** thì máy thực hiện lệnh lọc, nhưng các lệnh này không được lưu lại.

Sau khi lọc, SPSS tự động tạo ra biến lọc – biến “filter”, hiện thị ở dòng cuối xem trong **Variable view**

#### IV. Lập biểu tổng hợp

Chú ý: Mẫu 3% trong TĐTDS và nhà ở 1999 được phân bổ tương đối đều nhau ở các tỉnh/ thành phố. Do vậy kết quả chỉ có ý nghĩa sử dụng sau khi đã suy rộng số liệu. Trước khi lập các biểu tổng hợp cho các cấp, toàn bộ số liệu phải được suy rộng theo các quyền số W bằng cách:

Vào **Data**=>**Weight case**=>**Weight case by:**

- Đối với Data file: “người” suy rộng theo “W-người”
- Đối với Data file: “nhà” suy rộng theo “Weigh”
- Đối với Data file: “chết” suy rộng theo “Weigh”

Sau khi màn hình góc dưới bên phải hiện **“Weight on”** thì bắt đầu tổng hợp biểu.

Các lệnh thường dùng để xử lý tổng hợp kết quả:

Mở **Data file**=>**Analyze**=>**Descriptive Statistics**

- Thống kê {
1. Frequency (Tần xuất - tần số của biến)
  2. Descriptive (Mô tả biến)
  3. Explore (Thăm dò)
  4. Crosstabs (Tổng hợp chéo biến)

Trong các lệnh trên, thường dùng nhất là lệnh Crosstabs để tổng hợp biểu tương quan giữa các biến.

Vào **Crosstabs** màn hình hiển thị góc bên trái là các biến, góc bên phải là các biến sẽ được tổng hợp chéo theo dòng **Row** và theo cột **Column**. Phần **Layer** cho phép tổng hợp chi tiết phân lớp/ phân tầng; ví dụ: phân theo thành thị/nông thôn; phân theo vùng địa lý ..

Sau khi chuyển các biến cần xử lý theo dòng/ cột; bấm **paste** để lưu lệnh và bấm **run** hay bấm **OK** để chạy chương trình.

Để tổng hợp 1 biểu; SPSS chạy chương trình theo từng lệnh, từng Case. Đối với bản ghi “người” thông thường, sau khi chạy xong 2368167 Cases phải mất 10 phút sẽ cho ta một biểu tổng hợp. Cuối cùng màn hình sẽ hiển thị bên trái là các lệnh xử lý, bên phải là kết quả biểu tổng hợp. Từ đây ta có thể SAVE tên file có đuôi (\*. spo)

#### V. Chuyển các biểu tổng hợp sang bảng tính EXCEL

Bằng cách nhấn chuột vào màn hình trái, phần biểu tổng hợp của SPSS, dùng lệnh COPY, chuyển qua chương trình EXCEL và PASTE dưới dạng TEXT; Toàn bộ biểu tổng hợp sẽ được chuyển qua EXCEL. Trên bảng tính EXCEL, ta có thể trang trí lại tên biểu, khung biểu, làm các phép tính thông thường khác, có thể chuyển qua vẽ thành đồ thị dưới nhiều dạng khác nhau.