2- SPSS for windows

Cũng như Stata, SPSS đã từng qua các phiên bản chạy trên MS-DOS. Có thể chia SPSS thành 3 thế hệ chính như sau: các phiên bản trên MS- DOS; các phiên bản trên trước 8.0 Windows; các phiên bản từ 8.0 về sau trên Windows. Đặc điểm quan trọng của phần mềm này là ngay từ phiên bản trên MS-DOS, SPSS đã có hai cơ chế làm việc, đó là làm việc qua thực đơn và nhờ chương trình bằng ngôn ngữ riêng. Phần hoàn thiện chủ yếu tập trung vào hai mảng lớn là số liệu và đồ hoạ.

Về số liệu: SPSS chấp nhận hầu hết các format của các bảng tính, các hệ cơ sở dữ liệu và các phần mềm thống kê khác. Với DBMSCOPY 6.0, SPSS cho phép chuyển đổi tất cả các dạng số liệu lẫn nhau. Cấu trúc biến trong SPSS là một trong những cấu trúc hoàn thiện (kiểu biến, khuôn dạng, nhãn biến và giá trị nhãn). Trong quá trình xử lý, phân tích số liệu SPSS cho phép nối và chuyển đổi trực tiếp với các hệ cơ sở dữ liệu, các bảng tính. SPSS cho phép chia cắt các tệp số liệu cũng như ghép, nối các tệp số liệu cùng khuôn dạng, ngoài ra với các thủ tục riêng biệt việc lựa chọn tạm thời trong xử lý được tiến hành dễ dàng. Với modul Data entry, SPSS cho phép lập bảng hỏi và nhập số liệu trên khuôn dạng bảng hỏi mà không cần đến chương trình nhập số liệu riêng biệt, phức tạp như các cơ sở dữ liệu trước đây.

Về cấu trúc giao diện: SPSS cung cấp 4 màn hình (số liệu, kết quả, lệnh và khởi thảo báo cáo). Các màn hình này hoàn toàn có khả năng cắt dán lẫn nhau và cắt dán qua lại với các hệ soạn thảo, các bảng tính khác. Ngoài ra SPSS cho phép tạo chức năng chuyển đổi trực tiếp số liệu và khởi động Fox, Exel ... nhờ việc bổ sung ngay trong một thực đơn của SPSS các tệp chú của các phần mềm này. Trong trường hợp này ứng dụng được mở sẽ tự động đặt tên tệp là spss*.??? (phần mở rộng của ứng dụng tương ứng).

Về chương trình và các thủ tục: Hệ thống thực đơn luôn là ưu thế của SPSS. Có thể nói hầu như mọi công việc đều có thể thực hiện dễ dàng qua hệ thống thực đơn và bảng chọn. Bất kỳ một lựa chọn nào cũng sẵn sàng được giải thích "nóng" nếu ta kích chuột phải trên nó. Các thủ tục phức tạp nhất trong phân tích thống kê đã có mặt trên các thực đơn này (nhóm các thủ tục phân tích vô hướng, phân tích nhân tố và đặc biệt là phân tích chuỗi thời gian...). Song song với hệ thống thực đơn là ngôn ngữ SPSS, về phần này SPSS tỏ ra dễ tính trong cấu trúc ngữ pháp nhờ một hệ thống Autocorect. SPSS cũng cung cấp chương trình tương ứng với thủ tục được thực hiện cũng như các thông báo lỗi dưới hai dạng: Ngôn ngữ SPSS và ngôn ngữ viết trên Visual Basis hoặc Visual Java.

SPSS cũng cung cấp một hệ thống Macro cùng hệ song song Production facility. Chúng ta có thể tham khảo trong các tài liệu khác về chức năng này.

3- SPSS với mô tả thống kê

a- Mô tả thống kê với mẫu toàn bộ

Với một mẫu, mà cụ thể là một bộ số liệu SPSS cung cấp các mô tả thống kê bằng một số thủ tục cơ bản sau đây (Các chương trình kèm theo nhận được từ màn hình output):

+ Thủ tục Reports summaries in rows (analyse\Reports\summaries in rows)

frequencies the variables:

Number of Children	Ideal Number of Children	Age When First Married
Grand Total		
Mean 2	3	23
N 1495	965	1202
StdDev 2	2	5
Kurtosis 1.41	4.98	5.38
Variance 3	2	25
Skewness 1.04	2.24	1.66

+ Thủ tục frequencies: (analyse\descriptive statistic\frequencies)

```
FREQUENCIES
  VARIABLES=region4 married
  /NTILES=  4
  /STATISTICS=VARIANCE MEAN MEDIAN MODE SKEWNESS SESKEW KURTOSIS SEKURT
  /ORDER  ANALYSIS .
```

Thủ tục này cung cấp tất cả các đặc trưng thống kê mẫu cho một hay nhiều biến, một bảng mô tả chung và các bảng tần số cùng tần suất cho từng biến. Bảng tần suất có tính đến tần suất trong toàn bộ và tần suất trong số quan sát có thông tin. Sau đây là một phần kết quả của thủ tục này sau khi thực hiện lệnh các lựa chọn từ thực đơn tương ứng với lệnh trên.

Statistics

		Region	Married
N	Valid	757	1499
	Missing	743	1
Mean		2.5495	1.4696
Median		3.0000	1.0000
Mode		3.00	1.00
Variance		1.0098	.2492
Skewness		-.073	.122
Std. Error of Skewness		.089	.063
Kurtosis		-1.067	-1.988
Std. Error of Kurtosis		.177	.126
Percentiles	25	2.0000	1.0000
	50	3.0000	1.0000
	75	3.0000	2.0000

Region

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Northeast	136	9.1	18.0	18.0
	Midwest	221	14.7	29.2	47.2
	South	248	16.5	32.8	79.9
	West	152	10.1	20.1	100.0
	Total	757	50.5	100.0	
Missing	System	743	49.5		
Total		1500	100.0		

b- Mô tả thống kê với Select case và Split file

Thủ tục Select case cho phép chọn một bộ phận của mẫu theo một điều kiện nào đó hay chọn một cách ngẫu nhiên. Chúng ta cũng có thể chọn một bộ phận theo một cách như trên và xoá tất cả các bản ghi khác hay chỉ chọn tạm thời.

Sau đây là kết quả của một mô tả thống kê bộ phận với thủ tục Select case trên tệp GSS93.sav với các thủ tục được ghi lại trong chương trình tương ứng (số con của người goá).

```
    USE ALL.
    COMPUTE filter_$=(marital = 3).
    VARIABLE LABEL filter_$ 'marital = 3 (FILTER)'.
    VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
    FORMAT filter_$ (f1.0).
    FILTER BY filter_$.
EXECUTE .
    FREQUENCIES
      VARIABLES=childs
      /STATISTICS=STDDEV VARIANCE MEAN MEDIAN MODE
      /PIECHART  PERCENT
      /ORDER  ANALYSIS .
```

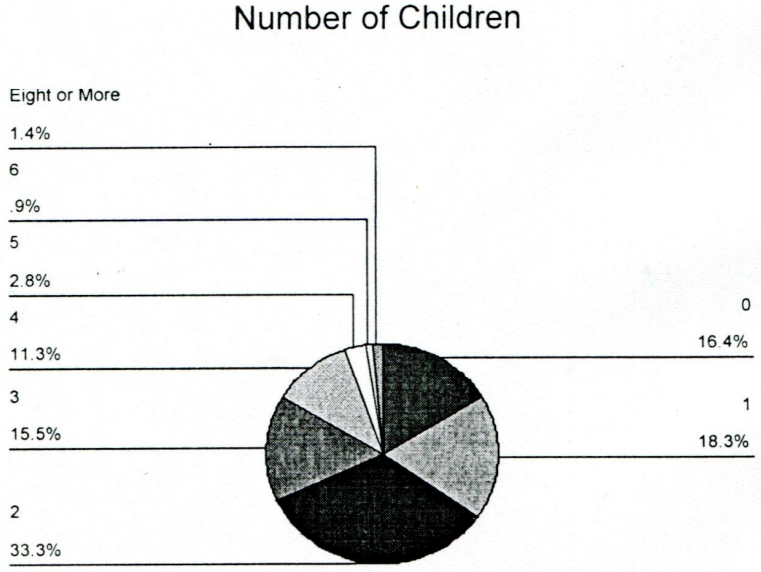
Statistics

Number of Children

N	Valid	213
	Missing	0
Mean		2.08
Median		2.00
Mode		2
Std. Deviation		1.54
Variance		2.36

Number of Children

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	35	16.4	16.4	16.4
	1	39	18.3	18.3	34.7
	2	71	33.3	33.3	68.1
	3	33	15.5	15.5	83.6
	4	24	11.3	11.3	94.8
	5	6	2.8	2.8	97.7
	6	2	.9	.9	98.6
	Eight or More	3	1.4	1.4	100.0
	Total	213	100.0	100.0	



Thủ tục Split file cho phép tách toàn bộ số liệu theo một biến nào đó và thực hiện mỗi thủ tục sau đó cho từng bộ phận. Sau đây là một phần kết quả một mô tả thống kê về số con với những người goá nhưng được tách theo vùng:

Region of Interview = New England

Statistics[b]

Number of Children

N	Valid	2
	Missing	0
Mean		1.00
Median		1.00
Mode		0[a]
Std. Deviation		1.41
Variance		2.00

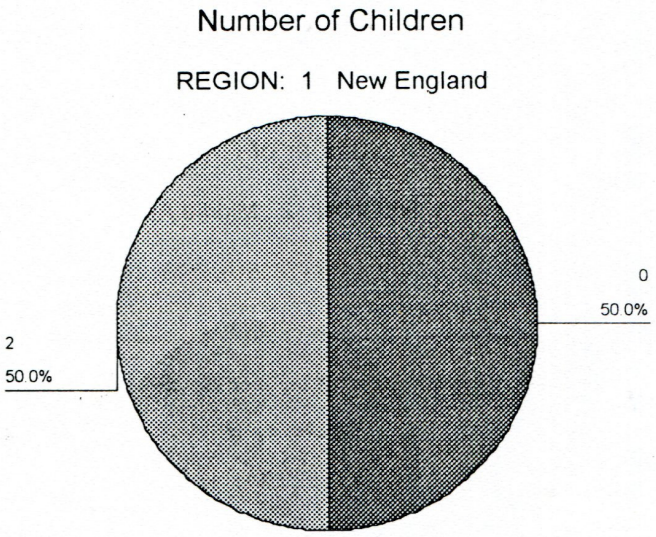
a. Multiple modes exist. The smallest value is shown

b. Region of Interview = New England

Number of Children [a]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1	50.0	50.0	50.0
	2	1	50.0	50.0	100.0
	Total	2	100.0	100.0	

a. Region of Interview = New England



Region of Interview = Middle Atlantic

Statistics[a]

Number of Children

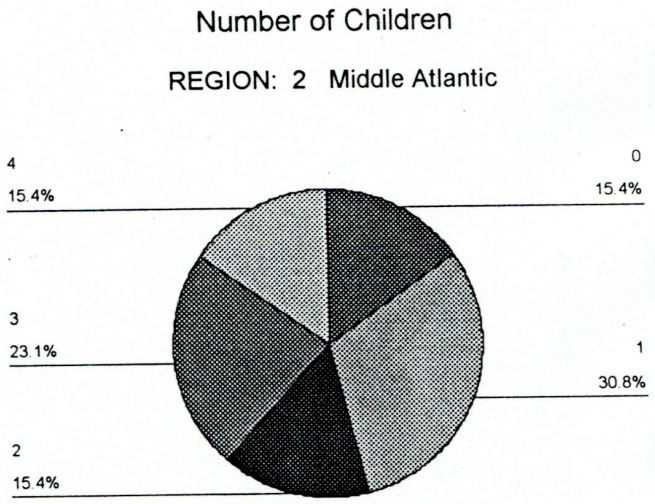
N	Valid	13
	Missing	0
Mean		1.92
Median		2.00
Mode		1
Std. Deviation		1.38
Variance		1.91

a. Region of Interview = Middle Atlantic

Number of Children[a]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2	15.4	15.4	15.4
	1	4	30.8	30.8	46.2
	2	2	15.4	15.4	61.5
	3	3	23.1	23.1	84.6
	4	2	15.4	15.4	100.0
	Total	13	100.0	100.0	

a. Region of Interview = Middle Atlantic

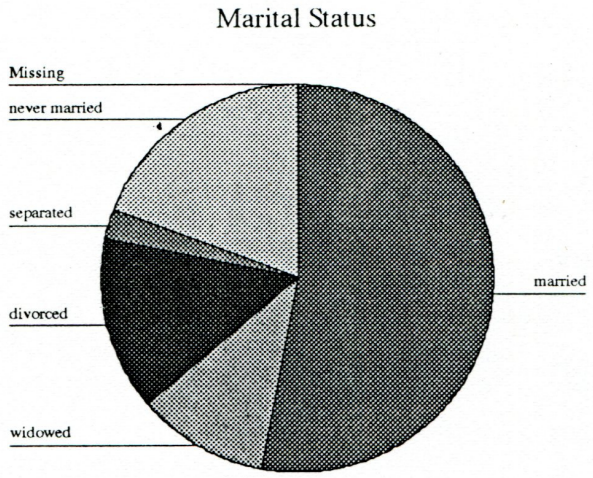


Với chương trình:

```
SORT CASES BY region .
SPLIT FILE
     SEPARATE BY region .
FREQUENCIES
 VARIABLES=childs
     /STATISTICS=STDDEV VARIANCE MEAN MEDIAN MODE
/PIECHART  PERCENT
     /ORDER  ANALYSIS .
```

Một chức năng khá mạnh của SPSS với biểu đồ là bộ phân hoá ngay trên biểu đồ. Chẳng hạn chúng ta có thể quan tâm đến biểu đồ của 1 biến, sau đó loại đi những phần không muốn hiển thị, SPSS cho phép làm lại trực tiếp biểu đồ với bộ phận được chọn.

Thí dụ: Biểu đồ % tình trạng hôn nhân từ tệp GSS93.sav

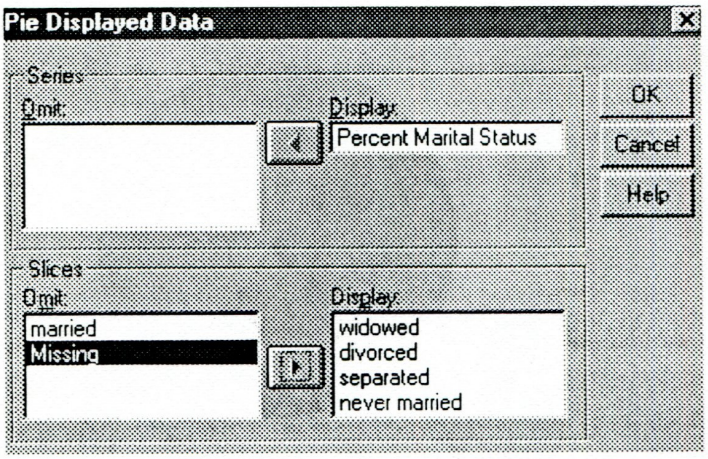


Biểu đồ này nhận được từ lệnh: FREQUENCIES

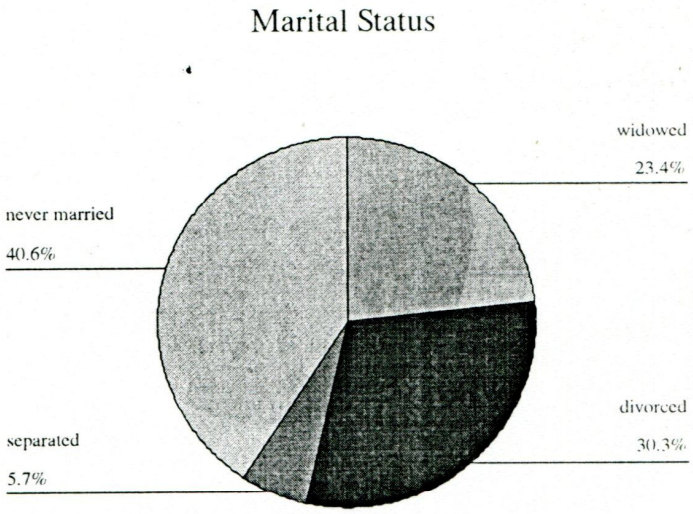
```
VARIABLES=marital  /PIECHART  PERCENT  /ORDER  ANALYSIS .
```

Chúng ta có thể loại bỏ hai nhóm: Không trả lời và hiện đang có vợ hoặc chồng ngay trên biểu đồ này nhờ kích hoạt biểu đồ hiện có và lựa chọn như sau:

Series\Displayed

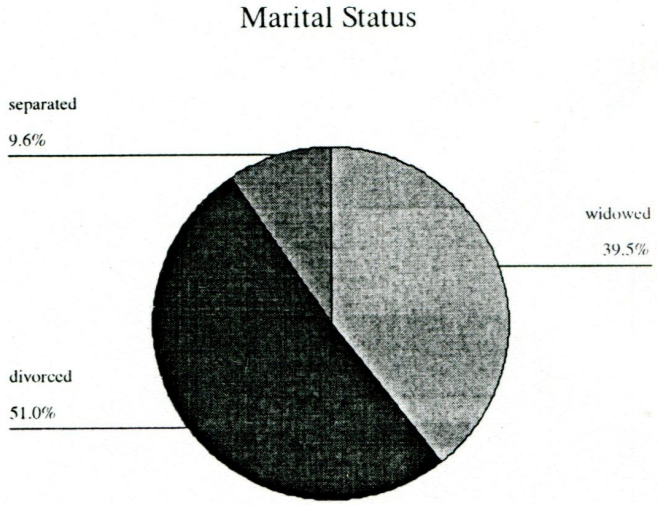


Kết quả là:



Các tỷ lệ đã được tính lại trong phạm vi 4 dấu hiệu.

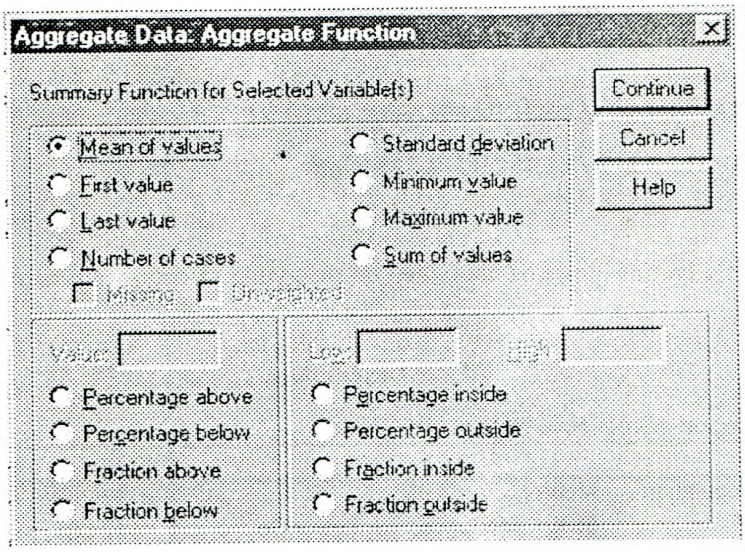
Nếu chúng ta muốn có hình ảnh về cơ cấu đã từng có vợ (hoặc chồng) nay không có vợ (hoặc chồng) thì có thể bỏ phần never married. Ta có biểu đồ sau:



c- Tạo tệp mô tả cơ bản với thủ tục Aggregate và các hàm

Cùng với các mô tả nhờ các đặc trưng thống kê, các biểu đồ, SPSS cung cấp thủ tục tóm tắt số liệu theo nhóm với một số hàm là các thống kê thông dụng.

Với Aggregate file chúng ta có thể chọn một hay nhiều biến phân loại và tạo tệp tóm tắt số liệu với các hàm trong bảng chọn sau:



Một tệp số liệu tóm tắt được tạo nhờ các thông tin có trong tệp số liệu hiện tại.

Hãy quan sát đoạn chương trình sau:

```
AGGREGATE
  /OUTFILE='C:\Program Files\SPSS\w95agg.sav'        {tệp mới có tên w95gg)
  /BREAK=region  /popula_1 = MEAN(populatn) /densit_1 = MEAN(density) /urban_1 = MEAN(urban)
  /lifeex_1 = MEAN(lifeexpf) /lifeex_2 = MEAN(lifeexpm) /litera_1 = MEAN(literacy) /pop_in_1 =
MEAN(pop_incr) /babymo_1 = MEAN(babymort) /gdp_ca_1 = MEAN(gdp_cap) /calori_1 = MEAN(calories)
/birth__1 = MEAN(birth_rt) /death__1 = MEAN(death_rt) /b_to_d_1 = MEAN(b_to_d)
  /fertil_1 = MEAN(fertilty) /lit_ma_1 = MEAN(lit_male) /lit_fe_1 = MEAN(lit_fema) /popula_2 =
PGT(populatn 100000) /litera_2 = PGT(literacy 70).
```

Không có gì khó khăn khi chép chương trình này vào màn hình Syntax và cho SPSS thực hiện chương trình. Bạn sẽ nhận được một tệp mới với các biến được lập từ tệp số liệu hiện có.

3- Thủ tục explore

Thủ tục này thực chất là lệnh Examine, nó cho phép mô tả các thống kê tóm tắt về một số biến theo một số biến khác. Các thống kê dùng kiểm định các giả thiết về tính đồng phương sai, tính chuẩn của phân phối cũng như các ước lượng khác nhau bằng phương pháp hợp lý tối đa. Các biểu đồ về các biến và các cặp biến. Có thể xem kỹ hơn thủ tục này ở phần lệnh Examine. Tuy nhiên thủ tục này cung cấp một số kỹ thuật quan trọng cho việc xác định giá trị trung tâm của các biến định lượng. Trước khi xem xét kỹ hơn các vấn đề trên ta quan sát sơ bộ thủ tục này.

Hãy chọn:

```
EXAMINE
  VARIABLES=populatn log_pop BY region /ID= country
```

```
      /PLOT    SPREADLEVEL    /MESTIMATORS    HUBER(1.339)  ANDREW(1.34)
HAMPEL(1.7,3.4,8.5)      TUKEY(4.685)/PERCENTILES(5,10,25,50,75,90,95)
HAVERAGE  /STATISTICS DESCRIPTIVES EXTREME
      /CINTERVAL 95  /MISSING LISTWISE  /NOTOTAL.
```

Sử dụng Pivot \Pivoting Trays để chọn hiển thị theo vùng và chọn vùng Latn America ta có bảng mô tả sau:

Descriptives

Region or economic group: Latn America

			Statistic	Std. Error
Population in thousands	Mean		21928.86	8031.90
	95% Confidence Interval for Mean	Lower Bound	5174.61	
		Upper Bound	38683.10	
	5% Trimmed Mean		15816.40	
	Median		7900.00	
	Variance		1.4E+09	
	Std. Deviation		36806.78	
	Minimum		256	
	Maximum		156600	
	Range		156344	
	Interquartile Range		17475.00	
	Skewness		3.020	.501
	Kurtosis		9.516	.972
Log (base 10) of Population	Mean		3.9738	.1286
	95% Confidence Interval for Mean	Lower Bound	3.7055	
		Upper Bound	4.2421	
	5% Trimmed Mean		3.9909	
	Median		3.8976	
	Variance		.347	
	Std. Deviation		.5894	
	Minimum		2.41	
	Maximum		5.19	
	Range		2.79	
	Interquartile Range		.6795	
	Skewness		-.280	.501
	Kurtosis		1.892	.972

Thủ tục này cung cấp một số ước lượng giá trị trung tâm của các biến qua lựa chọn M-estimation. (Xem kết quả ở mục II. 3)