# PERFORMING MULTIPLE CORRESPONDENCE ANALYSIS (MCA) USING SPAD<sup>1</sup> (VERSION 6.5)

Brigitte le Roux<sup>2</sup>, Mikael Börjesson<sup>3</sup>, Philippe Bonnet<sup>4</sup>

1. G	ENERALITIES	
1.1.	Main Window	1
1.2.	Choice of options	1
1.3.	Adding chains to "favorite chains"	2
2. IN	MPORTING DATABASES with DataXchange	
2.1.		
2.2.	Importing an SPSS database	6
2.3.	$\mathbf{r}$	
3. ST	TANDARD MCA	
3.1.	Performing a Standard MCA	
3.2.		
3.3.	Eigenvalues and modified rates	14
3.4.	$\partial$	
3.5.		
•	Choosing the preferences for graphs	
•	• Construction of graph for interpreting an axis	
3.6.		
4. S	TORING PRINCIPAL COORDINATES AND PARTITIONS	
5. A	PPENDIX	
5.1.	Generalities on the graph editor	
•	The toolbar of the graph editor	
•	The fundamental rule for formating a graph	
5.2.	Interface SPAD/SPSS using SPAD editor	
•	To Import SPSS databases: *.sav (SPSS) $\rightarrow$ *.sba (SPAD)	
•	To Export SBA databases: *.sba (SPAD) $\rightarrow$ *.sav (SPSS)	
•	Rules for converting SPAD to SPSS	

<sup>1</sup> SPAD WEB address : <u>www.SPAD.eu</u> Contact: Plilippe Pleuvret: <u>ppleuvret@spad.eu</u>

<sup>&</sup>lt;sup>2</sup> Université René Descartes PARIS. <u>www.math-info.univ-paris5.fr/~lerb</u> e-mail: <u>Brigitte.LeRoux@math-info.univ-paris5.fr</u>

<sup>&</sup>lt;sup>3</sup> Uppsala University – e-mail: <u>mikael.borjesson@ilu.uu.se</u>

<sup>&</sup>lt;sup>4</sup> Université René Descartes PARIS. e-mail: <u>Philippe.Bonnet@univ-paris5.fr</u>

# **1. GENERALITIES**

#### 1.1.Main Window

The **main window** of SPAD (which opens when you start the program) is composed of three main elements:

- 1. the toolbar with six menus (Dataset, Chain, Tools, Options, Window and Help),
- 2. the *Methods* window with 9 groups of methods (rolling menu): descriptive statistics, factorial analysis, etc.,
- 3. the *Chain* window which manages the linked sequence of methods applied to the chosen database that form a <u>chain</u>.<sup>5</sup>

SPAD		<u> </u>
Dataset Chain Tools Options Window !		
Methods	🔀 Chain 1:(no name)	
Descriptive Statistics	<u>C</u> hain <u>T</u> emplate <u>M</u> ethod <u>W</u> indow <u>H</u> elp	
(method name)		
	DATASET (no name)	
•		
08/09/2006		1.

## **1.2.** Choice of options

- 1. *Create* the directories <u>databases</u> and <u>chains</u> in your own working directory (for example in C:/.../My Documents/SPAD/).
- 2. Open the menu Options and choose General parameters.

SPAD				
Dataset Chain Tools	Options	Help		
Methods	<ul> <li>Show</li> </ul>			
Descriptive Statisti	<ul> <li>Show favourites chains</li> </ul>			
(method i	General parameters			
(method des	External applications			

3. The window <u>General parameters</u> opens.

```
Extension of a SPAD database:*.sbaExtension of a SPAD chain:*.fil
```

 $<sup>^{5}</sup>$  A SPAD chain is a graphical representation of the computations to be performed. At the top of a chain is the <u>Base icon</u> representing the database that SPAD uses for computation. The <u>Base icon</u> is followed by the <u>Method icons</u> that represent the requested computations. After the programming of methods and execution of the chains, the results will appear as icons on the right of the method icons.

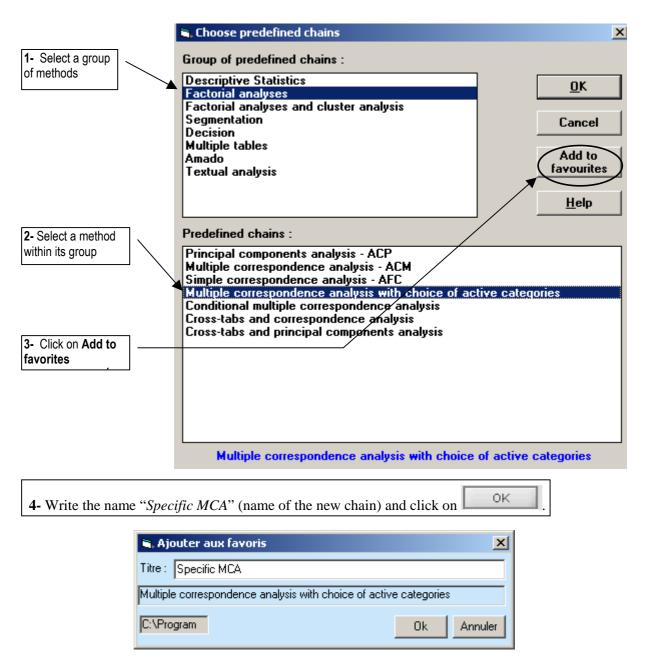
🐃 General parameters		
_ Default display ——		<u>o</u> ĸ
	Display of an empty chain	<u></u>
Start:	✓ Display of methods	Cancel
	<ul> <li>Display favorite chains</li> </ul>	Help
-Working parameters		
For each new chai	n : 💿 Open new dataset	
	O Re-open the last op	ened dataset
Packaging results	: 🗹 Automatic speeding	up Excel macros
Default directory for		
Datasets SPAD ·	C:\Program Files\spad\Spad_v60\data	sets
Templates :	C:\Program Files\spad\Spad_v60\Stan	d
Chains :	C:\Program Files\spad\Spad_v60\chai	15
Temporary zone:	C:\Program Files\spad\Spad_v60\Tmp	···
Imports :	C:\Program Files\spad\Spad_v60\Impo	rt
mode debug		/
a) Choose as <u>defat</u> directory, see a	ult directory: Dataset SPAD and Chains the bove 1.1.	ose created in your working

b) Click on

# 1.3. Adding chains to "favorite chains"

To add Specific MCA to favorite chains:

- open the <u>Chain</u> menu in the toolbar;
- choose <u>Predefined chains</u> and proceed as indicated hereafter.



Do the same for:

*"Standard MCA"* (Factorial Analyses <u>and</u> Multiple Correspondence Analysis). *"Standard MCA + AHC"* (Factorial Analyses and Cluster Analysis <u>and</u> Multiple Correspondence Analysis)

*"Specific MCA + AHC"* (Factorial Analyses and Cluster Analysis <u>and</u> Multiple Correspondence Analysis with Choice of Active Categories).

Hence the new window Favorite Chains:

×	Favorit chains
	Marginal distributions, histograms
<b>2</b>	Automatic characterisation of a nominal va
<b>2</b>	Automatic characterisation of a continuous
<b>2</b>	Principal components analysis
2	Multiple correspondence analysis
<b>2</b>	Interactive clustering tree
2	Interactive decision tree
2	Interactive regression tree
<b>2</b>	Discriminant analysis on categorical variab
Á	Standard MCA
Ŷ	Standard MCA + AHC
Ŷ	Specific MCA
Ŷ	Specific MCA + AHC

# 2. IMPORTING DATABASES with DataXchange

We will present the importation of an SPSS database (\*.sav).

**Before importation**, create a directory; name the directory (for example "SPSSbases"), then put in this directory the SPSS data files (\*.sav) that you want to import.

#### 2.1. New types of databases

1. Click on Dataset/Import/Databases

SPAD	
Dataset Chain Tools Option	ns Help
New dataset Edit dataset Ctrl+T Edit label	
Import 🕨	Import Ascii file
Export •	Import SAS Import textual data
Delete dataset	Databases
Exit	
	-

Then you enter the DataXchange module.

- 2. Choose <u>Preferences</u>, the second button with the right, click on Look&Feel and mark <u>Display quick buttons for mouse popup menu</u> and click on OK.
- 3. Click on the button <u>data source and destination configuration</u> at the center of the screen. The following window of configuration of <u>data source and destination</u> appears:

		L			
Velcome					
🍫 data	sourc	e and destination configu	ration	a 8	3
Source	s			b 🕼 🗈 🛷 💼 💿 🥝 ?¥	Ϋ́
visible	valid	name	type	description	
۲	?	my SBA data	sba	d new source.	
۲	?	my text files	Texte	text files directory	
۲	?	my Access Database	Access	Odbc link to SOCIO.mdb	
۲	?	my Excel File	xus	Odbc Excel Link	
Destina	tions				ation
Destina		name	E 22	i 🖻 of 💼 💿 🧭 ?¥	ation
Destina visible	valid	name my SBA data	type sba		ation
visible	valid ?	·/	type	ک 😰 🗈 💰 💼 👁 🧟 ?خ description	ation
visible	valid ? ?	my SBA data	type sba	A Image of the second	ation
visible @	valid ? ? ?	my SBA data my text files	sba Texte	A ■	ation
visible @ @	valid ? ? ?	my SBA data my text files my Access Database	sba Texte Access	Image: Second secon	ation
visible @ @	valid ? ? ?	my SBA data my text files my Access Database	sba Texte Access	Image: Second secon	ation

- 4. Use the button 🗎 (add new source) at the top of the window, and a new window will open, which will permit to specify what type of database to import
- 5. Choose <u>SPSS files</u> in the rolling menu:

select your connection type SPSS files OK Cancel	data source a	and desti	nation configuration
	select your c	connectio	n type
OK Cancel	SPSS files		-
		0K	Cancel

- 6. Then, click on
- 7. Write in the <u>field name</u>: "SPSS data"
- 8. and in the field description: "SPSS data"

0K

- 9. and then search for the folder ("SPSSbases", see above) where the SPSS dataset you want to analyze is located
- 10. click on and a new window will open (see window on the following page)
- 11. Double click, in the window labeled <u>Destinations</u>, on <u>my SBA data</u> to choose the directory in your working directory where you want to place the SPAD database and click on

Source	З		🔊 🖹 🍃	🖻 of 💼 👁 🧭 ?¥	
visible	valid	name	type	description	
۲	?	my SBA data	sba	dataset directory	-
۲	?	my text files	Texte	text files directory	
۲	?	my Access Database	Access	Odbc link to SOCIO.mdb	
۲	?	my Excel File	XLS	Odbc Excel Link	
۲	1	SPSS databases	SPSS	SPSS databases	
Destina	tions	$\sim$	🔪 🔒 🖹	🗈 🛷 🏨 🐵 💋 ?¥	
visible	valid	name	type	description	
۲	1	my SBA data	SBA	dataset directory	-
۲	?	my text files	Texte	text files directory	
۲	?	my Access Database	Access	Odbc link to SOCIO.mdb	
۲	?	my Excel File	XLS	Odbc Excel Link	

## 2.2. Importing a SPSS database

After defining the source in the previous step under the window <u>source</u>, open <u>SPAD DataXchange</u> again, click on <u>Begin new import/export project</u>. You then obtain the window below:

- 1. Window <u>Sources</u>: double-click on <u>SPSS databases</u> to select the database to import (culture.sav);
- 2. Window <u>Destinations</u>: double-click on <u>my SBA data</u> and thereafter on <u>new</u> to give a name (culture) to the SPAD database (with the extension \*.SBA: thus culture.SBA).

Sources		🌗 🕅	Destinations			🇳 🗰
	type				type	
🖶 Sources :			Destinations :	×		
🔍 🔍 my SBA data	sba	(	🖻 🗁 my SBA data	SBA		
💿 my text files	Texte	\ \	new 🖳 🕘			
🕞 🗁 SPSS databases	SPSS		inv text files	Texte		
Culture.sav						

3. Double click on the name of the SPSS database to import (culture.sav) to obtain a list of variables in the SPSS database.

4. Select the variables to export to SPAD;

•

- transfer the selected variables to the right window.
  - The type of the variables is **undefined**:
    - a) make an <u>automatic typing</u>, by selecting the variables and then click on the button

G Welcome	×2	Selecti	on 🖌		<b>3</b> <u>c</u>	onfiguration		4 Deploy	0
	type			$\overline{}$	new id	type	new name	description	
🕶 Sources :		+	🗏 Destina	tions :					
🖻 🔄 SPSS databases	\	Ē	🗄 🚮 cult	ure.sba					
🖻 🎹 culture.sav				21	1	undefined	Q1	do you prefer leisure activ	1
<b>@</b> Q1	numeric(label)			92	2	undefined	Q2	Would you say that during	J
<b>@</b> Q2	numeric(label)			93	3	undefined	Q3	If you had more time, 1st	]
<b>0</b> Q3	numeric(label) 🖊			24	4	undefined	Q4	When you go out in the ev	
<b>Q</b> Q4	numeric(label)			25	5	undefined	Q5	Time watching TV (hours	
<b>0</b> Q5	numeric(label)	I/		26	6	undefined	Q6	# of books or comic trips	
@ Q6	numeric(label)			281	7	undefined	QS1	Gender	
	numeric(label)			282	8	undefined	QS2	Education level	
QS2	numeric(label)		<b>Q</b> (	283	9	undefined	053	Age	
	numeric(label)		0 (	284	10	undefined	QS4	PCS	
QS4	numeric(label)	1		Z	11	undefined	z	one	
Z	numeric			D_SPADN	12	undefined	ID SPADN	SPAD identifier	
ID SPADN	string			15. N. N.			-		

Each variable is associated with one of the four following types<sup>6</sup>:

libel (identifying variable)<sup>7</sup> <u>nominal</u> (*categorical variable*) text (text variable) continu (numerical variable)

b) To change the type for a variable, double-click on the type for this variable and a new window appears that allows modifying the type of the variable (change the variable Z to nominal and id\_spadn to libel).

	type				new id	l type	new name	description
🕂 Sources :	100		📲 Destin	ations :			-	Setter Aller
🗄 🙆 SPSS databases			🗄 🚮 cu	lture.sba		<b>\</b>		
🗄 🧰 culture.sav			0	Q1	1	nominal	Q1	do you prefer leisure a
<b>9</b> Q1	numeric(label)		0	Q2	2	nominal	Q2	Would you say that dur
<b>@</b> Q2	numeric(label)		0	Q3	3	nominal	Q3	If you had more time, 1
<b>@</b> Q3	numeric(label)		0	Q4	4	nominal	Q.4	When you go out in the
<b>@</b> Q4	numeric(label)		- 0	Q5	5	nominal		Time watching TV (hor
<mark>0</mark> Q5	numeric(label)	1000	0	Q6	6	libel	Q6	# of books or comic trip
<b>9</b> Q6	numeric(label)		0	QS1	7	nominal	QS1	Gender
<mark>0</mark> QS1	numeric(label)		0	QS2	8	texte	QS2	Education level
	numeric(label)			QS3	9	continu	QS3	Age
<mark>0</mark> Q83	numeric(label)		0	QS4	10	nominal	QS4	PCS
	numeric(label)	1	9	Z	11	continu	z	one
Z	numeric			ID_SPADN	12	nominal	ID_SPADN	SPAD identifier
ID_SPADN	string							

5. For ordinal variables, verify that the order of the modalities is  $correct^8$ .

7/8

<sup>&</sup>lt;sup>6</sup> The types of variables are in French: nominal=categorized; libel=label; continu=numerical. <sup>7</sup> There can only be one identifying variable.

<sup>&</sup>lt;sup>8</sup> In the last version of DataXchange (October 2006), modalities are listed according to the order of the SPSS file. In the preceding version, the import is made according to the alphabetical order of the labels of the modalities (categories).

41(1)	ick on enter categorie	25	SPAD Da	ataXchange	
4.1. Ch	erer on <u>enter categoria</u>	<u> </u>		utau 🛛	
		$\backslash$		ndex 9	
		$\backslash$	t	ype 1	nominal 💌
		$\sim$	n	ame 🛛	253
			des	cription	\ge
				variable i	s recoded
				Ø missin	g values
	<i>4.2a</i> . Cli	ck on <u>ok.</u>	×	enter cate	egories
	4.2b. Clic	k on <u>no</u>		canc	el ok
		SPAD DataXchang	e		
		do vou want to	keep using the (	existing ca	tegories list ?
4.3. Sel	lect search all		heep doing hee		logonoo not :
categorie	es (maximum)				
then clic	ck on <u>0k.</u>		yes	no	cancel
			,		
1	SPAD DataXchange 🛞				
	) search categor	ies on the		100 📜 fir	stlines
	<ul> <li>search all cate</li> </ul>				stinct values)
	Search an cate	gones (maximum			
				cance	el ok
				cance	el <b>C</b> ok
	Modalities			cance	el ok
	Modalities Modalities	1	<b>~</b>	cance	
4.4 to change	order <u> manua</u>	L	 label		Z ↓ªZ №
<i>4.4.</i> to change the order,		I short name m0	▼  abel <=25		Z <sup>7</sup> X x ↓ <sup>2</sup> Z % values
the order, choose the	order 🜗 manua	short name			values
the order,	order 🖑 manua index 1	short name m0	<=25	F [<=25	values ] 5]
the order, choose the	order <table-cell-columns> manua index 1 2 3 4</table-cell-columns>	m0 m1 m2 m3	<=25 35-45 55-65 25-35	[<=25 [35-4: [55-6: [25-3:	values 3 5 5 5 5 5 5 5 5 5 5 5 5 5
the order, choose the	order index 1 2 3 4 5	short name m0 m1 m2 m3 m4	<=25 35-45 55-65 25-35 45-55	[<=25] [35-4: [55-6: [25-3: [45-5:	values 3 5 5 5 5 5 5 5 5 5 5 5 5 5
the order, choose the	order <table-cell-columns> manua index 1 2 3 4</table-cell-columns>	m0 m1 m2 m3	<=25 35-45 55-65 25-35	[<=25 [35-4: [55-6: [25-3:	values 3 5 5 5 5 5 5 5 5 5 5 5 5 5
the order, choose the	order index 1 2 3 4 5 6	short name m0 m1 m2 m3 m4	<=25 35-45 55-65 25-35 45-55	[<=25] [35-4: [55-6: [25-3: [45-5:	values values 5] 5] 5] 5]
the order, choose the	order index 1 2 3 4 5 6	short name m0 m1 m2 m3 m4 m5	<=25 35-45 55-65 25-35 45-55	[<=25 [35-4] [55-6] [25-3] [45-5] [>65]	values yalues 5] 5] 5] 5]

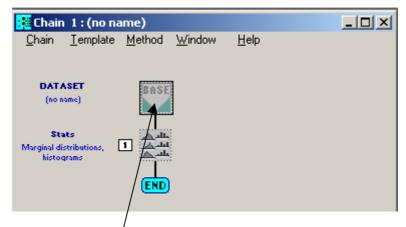
• For example, for age, double-click on QS3, and the following window appears

- 6. Click on the button <u>Deploy</u>, you will return to the initial window.
- 7. Click on the button Launch the import/export, then click on No, and the database culture.sba will be created in the folder <u>Bases</u> in your working directory.

## 2.3. Checking of the imported SPAD database (sba)

Check that the imported data corresponds to the initial SPSS database, for instance by comparing the frequencies of the two files.

Choose <u>Marginal distributions, histograms</u> in the window <u>Favorite chains</u>. You obtain the following window <u>Chain</u>:

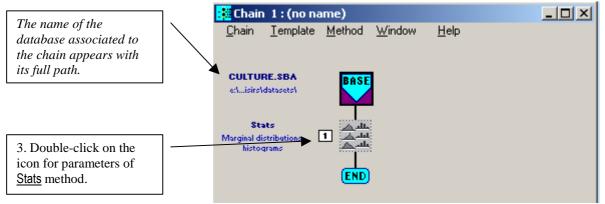


The icons are grey, which means that the methods are not parameterized (that is, the options are not specified) $^9$ .

To select the database culture.sba: /

- 1. double-click on the icon BASE
- 2. in the list of databases, double-click on culture.sba

You obtain:



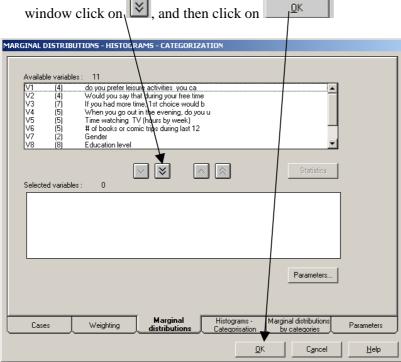
The parameter window is structured in different *sheets*, each sheet groups the different parameters of the method.

<sup>&</sup>lt;sup>9</sup> Colors of icons inform about their state:

grey: the method is not parametrized.

yellow: the method is parametrized.

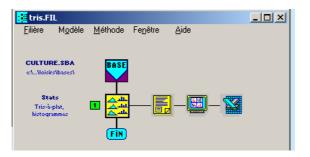
3. Click on the sheet Marginal distributions and in order to transfer all the variables to the button



**4.** Click on the icon <u>STATS</u> with the right button of the mouse and on <u>Execute</u> the method (it is first required to save the chain).



5. Results.



To obtain the frequency tables as a text file, click on the frequency tables as a text file, click on

# **3. STANDARD MCA**

#### 3.1. Performing a Standard MCA

Choose the chain "<u>Standard MCA</u>" in the window <u>Favorite chains</u>, and choose the database by doubleclicking on the icon and choose culture.sba.

You obtain the following window:

🔀 Chain 1 : (no na	ime)			_ 🗆 ×
<u>C</u> hain <u>L</u> emplate	Method	<u>W</u> indow	<u>H</u> elp	
CULTURE.SBA e:IisirsIdatasetsI				
Cormu Multiple correspondence analysis (ACM)	1	]		
Defac Description of factorial axes	2 .			
	END			

It is then necessary to set the parameters for the <u>Cormu</u> and <u>Defac</u> methods.

#### • Setting parameters for Cormu

Double-click on the icon . The window of parameters for the Standard MCA has 4 sheets: <u>Variables</u>, <u>Cases</u>, <u>Weighting</u>, <u>Parameters</u>.

Click on the sheet <u>Variables</u>, in order to select the active questions (variables) and the supplementary questions (variables).

1.a. Select the active questions (rolling menu: Variable selection: Active categorical variables): and

transfer the first six variables below by using the button with one arrow  $\square$ . 1.b. Do the same for the supplementary questions (Supplementary categorical variables).

TIPLE CORRESPONDEN	E ANALYSIS	_	/
Variables selection Available variables: 1	Active categorical variables Active categorical variables Supplementary categorical variable Supplementary continuous variable		
V2 (4) Wo V3 (7) Ifyo V4 (5) Wh V5 (5) Tim V6 (5) #ol	iou prefer leisure activities you ca uld you say that during your free time u had more time, 1st choice would b en you go out in the evening, do you e watching TV (hours by week) books or comic trips during last 12 der		
V8 (8) Edu Selected variables:			Statistics

2. Click on the sheet <u>Cases</u> to choose the active individuals (cases) and the supplementary individuals (cases). Choose All.

MULTIPLE CORRESPONDENCE ANALYSIS					
Choose cases	Sampling on chose	en cases			
O Logical filter O List O Interval	⊙ No O Yes	Define	Save	Use	

3. Click on the sheet <u>Weighting</u>.

MULTIPLE CORRESPONDE	NCE ANALYSIS	
O Weighting variable	V11 (CONT) one	<b>Y</b>
Colouistion by distribu	ition adjustment	

4. Click on the sheet Parameters.

Retained coordin 2 Random assignment of ac categories inferior to (	tive 0.000	GDA methodology, use s CA putting rare modalitie	
Printout parameters Multiple correspondence t ( 3 Printed coordin 4 Results for the c	able Burt) No 💌 Nates 💽 The first 3		t, the coordinates of
File for Excel applic	ation 🕑 Yes 🔹 Options	the individ in the outp No	duals are not included put.
			_

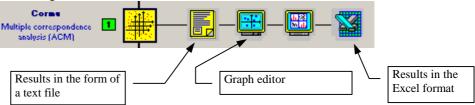
#### • Running MCA

Click on the right button of the mouse on the icon

and choose the Run method.

#### 3.2. Results of MCA

You obtain the following icons:



Verify the frequencies and the choice of active and supplementary questions and active individuals in the text file.

Printout	
	SELECTION OF CASES AND VARIABLES
🗄 🗊 Selection of cases and variables	ACTIVE CATEGORICAL VARIABLES 6 VARIABLES 30 ASSOCIATED CATEGORIES
in all the series of the seri	
	3. If you had more time, 1st choice would b     (7 CATEGORIES )       4. When you go out in the evening, do you u     (5 CATEGORIES )
	5. Time watching TV (hours by week) ( 5 CATEGORIES ) 6. # of books or comic trips during last 12 ( 5 CATEGORIES ) 
	SUPPLEMENTARY CATEGORICAL VARIABLES 4 VARIABLES 27 ASSOCIATED CATEGORIES
	7. Gender         ( 2 CATEGORIES )           8. Education level         ( 8 CATEGORIES )           9. Age         ( 6 CATEGORIES )           10. PCS         ( 11 CATEGORIES )
	CASES
	WEIGHT OF CASES         : Weight of objects, uniform equal to 1.         UNIF           KEPT
	SUPPLEMENTARY NISUP = 0 PISUP = 0.000
Printout ⊕-j⊒ Selection of cases and va ⊕-j⊒ Multiple correspondence	TOTAL WEIGHT OF ACTIVE CASES : 2720.00
🗄 🗊 Selection of cases and va	ELIMINATION OF ACTIVE CATEGORIES WITH SMALL WEIGHTS           ariables           ariables           GroBysit           Gnobysit           Libert           Label           LOBENT           LOBENT           LOBENT           LOBENT           Antipologic           Schement           Schement           Lobent           Lobent      <
🗄 🗊 Selection of cases and va	HLININATION OF ACTIVE CATEGORIES WITH BMALL WEIGHTS           ariables           ariables           STOBUST           BARDAR CLEANING           CATEGORIES           BEDRE CLEANING           CATEGORIES           BEDRE CLEANING           CATEGORIES           CATEGORIES           IDENT           LABLL           CONTEGORIES           Distriction           1           do you prefer leisure activities you ca mado: Leisurefriefied           1000           B000           B000           B000           1.1           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           1.2           2.2           2.2           2.2           2.2           2.2           2.2           2.2            2.2
🗄 🗊 Selection of cases and va	HLHNIARION OF ACTIVE CATEGORIES WITH SMALL WEIGHTS       0.80         ariables       1.83 % WEIGHT:       0.80 % WEIGHT:         ariables       6 ACTIVE QUESTIONS       30 ASSOCIATE CATEGORIES         ariables       10714.WEIGHT:       0.80 % WEIGHT:       0.80 % WEIGHT:         ariables       10714.WEIGHT:       0 ASSOCIATE CATEGORIES       0 ASSOCIATE CATEGORIES         ariables       10714.WEIGHT:       0 ASSOCIATE CATEGORIES       0 ASSOCIATE CATEGORIES         ariables       1.40 you prefer leisure activities you ca       0 PTER CLEANING       AFTER CLEANING         1. do you prefer leisure activities you ca       1.80 1861.01 100 PTER CLEANING       0 PTER CLEANING         1. do you prefer leisure activities you ca       1.80 1861.01 100 PTER CLEANING       0 PTER CLEANING         1. do you prefer leisure activities you ca       1.80 1861.01 100 PTER CLEANING       0 PTER CLEANING         1. do you prefer leisure activities you ca       1.80 1861.01 100 PTER CLEANING       0 PTER CLEANING         1. do you prefer leisure activities you ca       1.80 1861.00 PTER CLEANING       0 PTER CLEANING         1. do you say that during your free time       1.81 597 597.00 PTER CLEANING       0 PTER CLEANING         1. lack time       1.121 101.10 PTER CLEANING       0 PTER CLEANING       0 PTER CLEANING         2. Would you say that during your
🗄 🗊 Selection of cases and va	HLININATION OF ACTIVE CATEGORIES WITH BMALL WEIGHTS         aniables         aniables         SCONVEX         BASE         Conserved         Co
🗄 🗊 Selection of cases and va	<pre>Hi.HHINATION OF ACTIVE CATEGORIES WITH BMAIL WEIGHTS BERNOLD (FKNN): 1.03% WEIGHT: 0.88 BERNOLD (FKNN): 6 ACTIVE QUESTIONS 30 ASBOCIATE CATEGORIES BERNOLD (FKNN): 1.03% WEIGHT: 0.88 BABOCIATE CATEGORIES CATEGORIES : 277:10 MARGINAL DISTRIBUTIONS OF ACTIVE QUESTIONS CATEGORIES : DECOMPT WEIGHT: CATEGORIES MARGINAL DISTRIBUTIONS OF ACTIVE QUESTIONS CATEGORIES : DECOMPT WEIGHT: BUTTORS CATEGORIES : DECOMPT WEIGHT: BUTTORS DECOMPT WEIGHT: DISTRIBUTIONS OF ACTIVE QUESTIONS CATEGORIES : DECOMPT WEIGHT: BUTTORS DECOMPT WEIGHT: DISTRIBUTIONS OF ACTIVE QUESTIONS CATEGORIES : DECOMPT WEIGHT: BUTTORS DECOMPT WEIGHT: DISTRIBUTIONS OF ACTIVE QUESTIONS CATEGORIES : DECOMPT WEIGHT: BUTTORS DECOMPT WEIGHT: DISTRIBUTIONS OF ACTIVE QUESTIONS CATEGORIES : DECOMPT WEIGHT: BUTTORS DECOMPT WEIGHT: DISTRIBUTIONS OF ACTIVE QUESTIONS CATEGORIES : DECOMPT WEIGHT: BUTTORS OF ACTIVE DECOMPT WEIGHT: DISTRIBUTIONS OF ACTIVE QUESTIONS DECOMPT WEIGHT: DISTRIBUTIONS OF ACTIVE ACTIVE ACTIVE DECOMPT WEIGHT: DISTRIBUTIONS OF ACT</pre>
🗄 🗊 Selection of cases and va	HLININATION OF ACTIVE CATEGORIES WITH BMALL WILGHTS       1.00 % WILGHTS       1.00 % WILGHTS         aniables GROUVE       Intension (Contension)       0.00 % WILGHTS       10.00 % WILGHTS         aniables GROUVE       Intension (Contension)       0.00 % WILGHTS       10.00 % WILGHTS         aniables GROUVE       Intension (Contension)       0.00 % WILGHTS       10.00 % WILGHTS         GROUVE       Contension (Contension)       10.00 % WILGHTS       10.00 % WILGHTS         Intension (Contension)       Intension (Contension)       After Cleaning       10.00 % WILGHTS         Intension (Contension)       Intension)       Intension (Contension)       Intension)         Intension (Contension)       Intension)       Intension)       Intension (Con

The results in the Excel document are the following:

Cormu-1: marginal distributions of active variables

Cormu-4: control panel of eigenvalues

Cormu-5: loadings [coordinates] of active categories [modalities]

Cormu-6: contributions of active categories [modalities]

Cormu-7: squared cosines of active categories [modalities]

Cormu-8: loadings [coordinates] of active and supplementary categories [modalities]

Cormu-9: test-values of active and supplementary categories [modalities]

To interpret axes, we will essentially use the  $\underline{Cormu-4}$  and  $\underline{Cormu-6}$  sheets and then construct graphs of modalities for the interpretation of the axes.

#### **3.3.** Eigenvalues and modified rates

We use <u>Cormu-4</u> complemented with modified rates.

To calculate the modified rates, make the following calculations:

- 1) Modified values (column E) for the eigenvalues inferior to the average eigenvalue (that is 1/Q, where Q is the number of active variables, in this case: 1/6=0.1666),
- 2) Modified rates (modified values divided by the sum of all modified values specified in column E).
- 3) Cumulated modified rates.

	A	В	С	D	E	F	G	
1	Control par	el of Eigenv	alues			_ <b>F</b> 4	=E4/\$E\$28	
2	Trace of ma	atrix: 4.	00000	E4=(B4-1)	<b>(6)^2</b>			
							Cumulated	
	Number	Eigenvalue	Percentage	Cumulated	Modified	Modified	modified	
З		Ū	Ū	Percentage	values	rates	rates	
4	1	0.2925	7.31	7.31	0.01583982	0.5691748	0.5691748	
5	2	0.2415	6.04	13.35	0.00559928	0.20119985	0.77037465	
6	3	0.2248	5.62	18.97	0.0033839	0.12159439	0.89196904	$\Box$
7	4	0.2073	5.18	24.15	0.00164749	0.05919968	0.95116872	
8	5	0.1950	4.87	29.03	0.00080164	0.02880564	0.97997436	
9	6	0.1832	4.58	33.61	0.00027438	0.00985924	0.9898336	$\square$
10	7	0.1790	4.47	38.08	0.00015088	0.00542161	0.99525521	
11	8	0.1758	4.40	42.48	8.4187E-05	0.0030251	0.9982803	
12	9	0.1733	4.33	46.81	4.3367E-05	0.0015583	0.9998386	
13	10	0.1688	4.22	51.03	4.4916E-06	0.0001614	1	
14	11	0.1655	4.14	55.17				
15	12	0.1605	4.01	59.18				
16	13	0.1593	3.98	63.16		G4=SUN	I(\$F\$4:F4)	
17	14	0.1562	3.90	67.06				
18	15	0.1525	3.81	70.88				
19	16	0.1506	3.77	74.64				
20	17	0.1445	3.61	78.25				
21	18	0.1427	3.57	81.82	[ Si	um of modifie	d values of	
22	19	0.1363	3.41	85.23	ei	genvalues inf	erior to	
23	20	0.1309	3.27	88.50	1/	Q=1/6=0.166	6	
24	21	0.1249	3.12	91.62		28=SUM(E24:	E27)	
25	22	0.1203	3.01	94.63	(c	ell \$E\$28)		
26	23	0.1168	2.92	97.55				
27	24	0.0979	2.45	100.00	] 🖌			
28					0.02782944			
29								

#### **3.4.** Interpretating axes using contributions

The interpretation of the axes is based upon the contributions of the categories [mdalities] (given in sheet <u>cormu-6</u>). For each axis, one marks the categories [modalities] whose contributions are above average contribution, that is, 100/30=3.3% (here one has 30 active categories). See the following table, in which the most contributing modalities are highlighted.

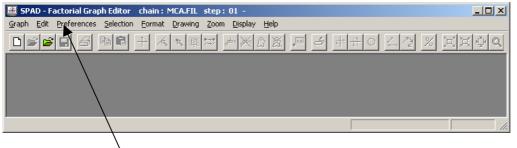
Contributions of active categories			COR	MU_6	
Label	Relative Weight (%)	Squared distance to origin	Axis 1	Axis 2	Axis 3
do you prefer leisure activities you ca					
Leisure:friends	6.495	1.56604	7.33	8.53	6.11
Leisure:family	4.406	2.78303	3.73	12.89	4.53
Leisure:alone	2.659	5.26728	2.66	4.58	9.27
Leisure:partner	3.107	4.36489	0.01	3.71	10.83
TOTAL	16.667		13.73	29.71	30.74
Would you say that during your free ti	me			•	
lack time	6.869	1.42640	3.61	4.47	0.12
always sth to do	7.09	1.35091	1.35	0.24	0.58
Stimes nothing to do	1.477	10.28630	0.03	3.06	1.96
often do nothing	1.232	12.53230	3.52	3.60	1.21
TOTAL	16.667		8.50	11.37	3.88
If you had more time, 1st choice would					
home DIY (do it yourself)	2.586	5.44550	4.51	0.01	0.85
artistic activities	2.414	5.90355	3.98	0.00	2.87
to rest	1.863	7.94737	4.53	4.24	0.06
develop knowledge	2.751	5.05791	0.33	0.33	7.39
physical activities	3.511	3.74695	4.60	0.00	4.43
take care of family	1.936	7.60759	2.16	9.29	1.50
to take courses	1.605	9.38168	0.05	0.16	4.32
TOTAL	16.667		20.17	14.03	21.41
When you go out in the evening, do you	uu			I.	
GoingOut:friends	3.627	3.59459	9.45	12.93	7.45
GoingOut:alone	1.238	12.46530	0.02	0.94	1.04
GoingOut:partner	5.582	1.98573	1.11	4.09	9.91
don't go out	2.874	4.79957	15.42	3.47	2.13
GoingOut:family	3.346	3.98169	1.00	11.89	10.24
TOTAL	16.667		27.00	33.32	30.78
Time watching TV (hours by week)					
TV:never	4.320	2.85816	0.03	0.91	0.04
TV:<10h	4.865	2.42569	1.10	1.17	0.23
TV:[10;19h]	3.241	4.14178	8.18	3.45	1.12
TV:[19;30h]	2.665	5.25287	3.59	0.15	0.89
TV:>=30h	1.575	9.58366	0.01	1.75	2.14
TOTAL	16.667		12.91	7.43	4.43
# of books or comic strips during last 1	2			•	
no book	3.928	3.24337	0.53	0.54	0.03
1-4 books	2.953	4.64315	0.67	1.29	1.32
5-12 books	3.695	3.51078	10.02	1.37	2.08
13 -39 books	2.641	5.31090	1.21	0.89	5.21
40 books or more	3.450	3.83126	5.25	0.04	0.14
TOTAL	16.667		17.69	4.13	8.78

## 3.5. Graph for Interpretating Axes

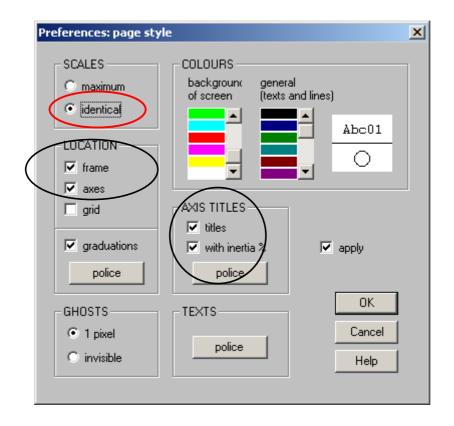
Choosing the preferences for graphs •

To enter the graph editor, double-click on the icon

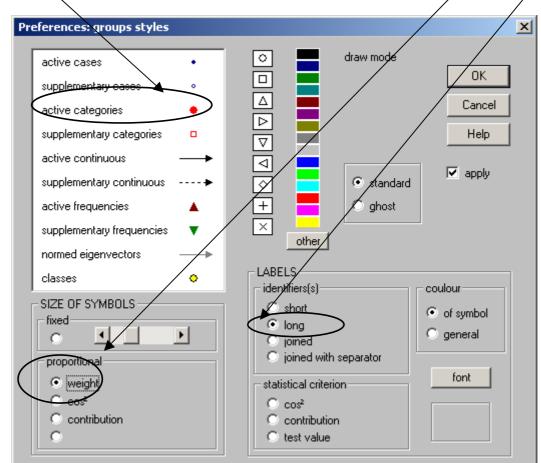


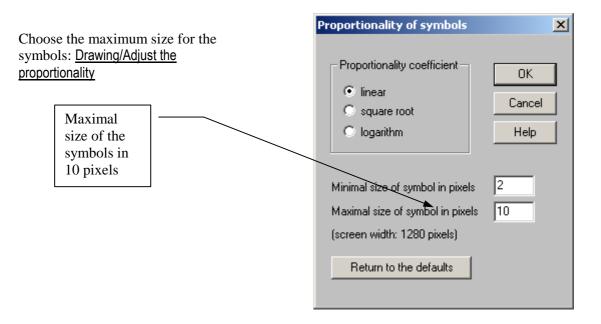


Define the <u>Preferences</u>/ Style for the page



Define the <u>Preferences</u> for the active categories (<u>style for the groups</u>), define one color and a symbol for the active category (for example a red filled circle), a size that is proportional to the weight and long labels as indicated below.





#### • Construction of graph for interpreting an axis

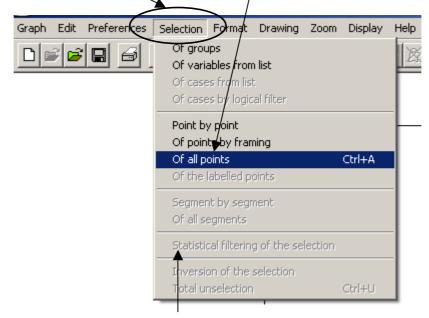
1. Select <u>Graph/New</u>, which gives you the following window:

Pre	eselections for a new graph		×
	active cases	✓ active frequencies	1
	supplementary cases	🔽 supplementary frequencies	
	active categorical variables	🔽 normed eigenvectors	
	<ul> <li>supplementary pategorical variables</li> <li>active continuous variables</li> </ul>		
	Supplementary continuous variables	cases draft variables selection	
	ОК	Cancel Help	

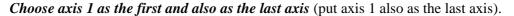
- 2. Select the active questions by marking active categorical variables.
- 3. If preferred, redraw the graph symmetrically to the horizontal axis, by using

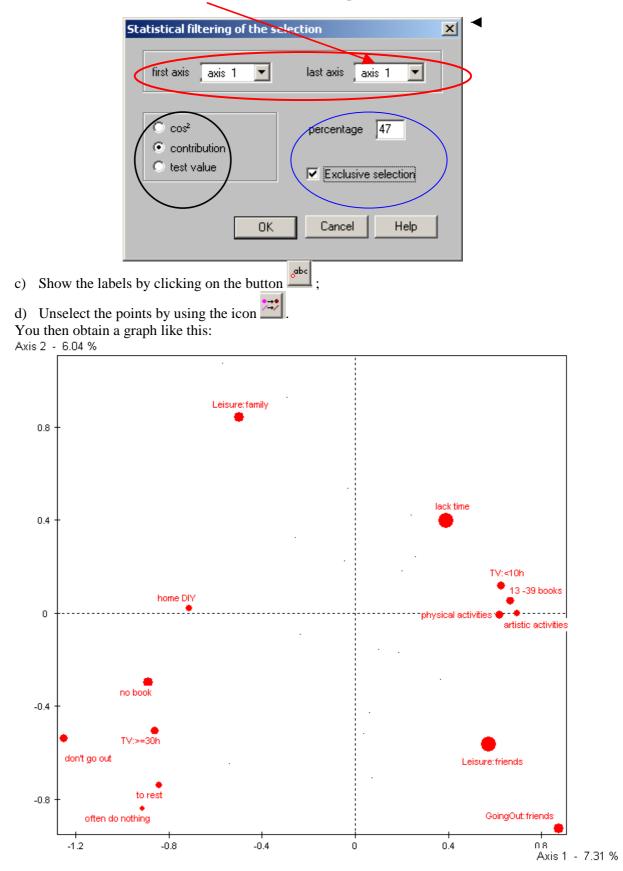


- 4. In order not to show more than 14 out of the 30 (46%) categories that contribute the most to axis 1:
  - a) Select all points (Selection/Of all points) (the selected points becomes purple)



b) Then Selection/Statistical filtering of the selection



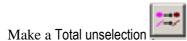


To interpret axis 2, perform the same steps as for axis 1: use contributions to axis 2 as the statistical criteria for selecting the modalities and drawing the graph. Do the same for axis 3.

## 3.6. Graph of the Cloud of Individuals

To obtain the cloud of individuals with point sizes proportional to superposition:

- 1. Parameters of proportionality: <u>Drawing/Adjust the proportionality</u> and choose <u>Maximal size of the</u> <u>symbols in pixels</u> (for example 8).
- 2. Select all the points (<u>Selection/Of all points</u>), and go to the menu <u>Format/Colours, symbols</u>,... and check the item <u>Proportional size: Superposition</u>.

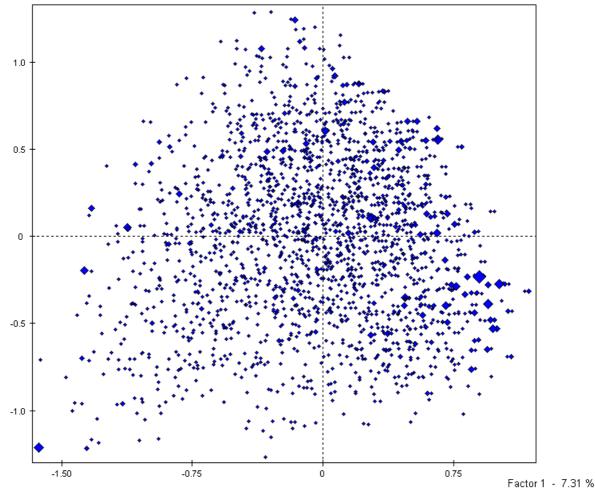


4. If you wish to redraw the graph symmetrically to the horizontal axis, use

You then obtain, in the plane of axes 1-2, the following graph:



3.



Let us now study the cloud of individuals structured by age.

	/		
Pre	eselections for a new graph		×
(	active cases	<ul> <li>active frequencies</li> <li>supplementary frequencies</li> </ul>	
	🔲 active categorical variables	🔽 normed eigenvectors	
(	<ul> <li>supplementary categorical variables</li> <li>active continuous variables</li> </ul>	I partitions	
	supplementary continuous variables	cases draft	
		variables selection	
	ОК	Cancel Help	

Select Graph/new: check Active cases and Supplementary categorical variables

and select the variable Age which will function as structuring factor, by clicking on variables selection.

(If you wish to redraw the graph symmetrically to the horizontal axis, use ), Parameters of proportionality: <u>Drawing/Adjust the proportionality</u> and choose <u>Maximal size of the symbols in</u> <u>pixels</u> (for example 6).

Choose <u>Selection/Of Groups/Supplementary Categories</u> and then Format/Colours, Symbols,...: Choose for example Red and empty squares.

Then Total unselection

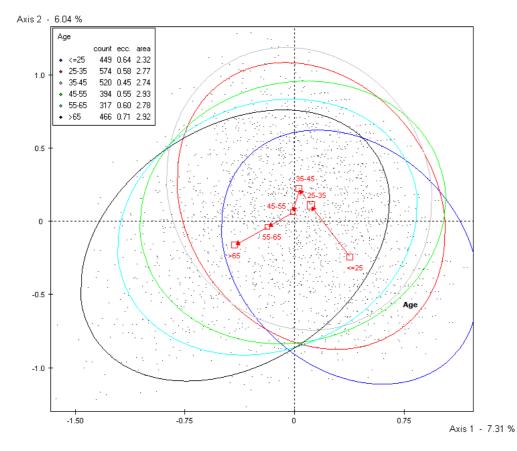
To join the categories of age, click on , choose Age, and click OK. To trace the 6 ellipses: Format/Cases by a categorical variable or a partition and click on Age...

mat of cases by a catego	rical variable or a partition	
active cases	© colours	ОК
	C character form	Cancel
categorical variables	return to the default format	Help
1 do you prefer leisure au 2 Would you say that du 3 If you had more time, 1 4 When you go out in the 5 Time watching TX (ho 6 # of books or comic tri 7 Gender 8 Education level 9 Age	ring your free time st choice would b e evening, do you u jurs by week)	

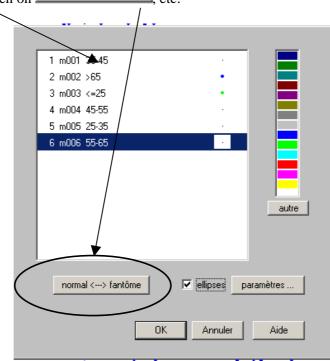
... and check ellipses in the following window:

1 m0 <=25	•			
2 m3 25-35	•			
3 m1 35-45	*			
4 m4 45-55	•			
5 m2 55-65	*			
6 m5 >65	•	other		
standard <> ghost	ellipses paramet	ers		
	OK Cancel H	lelp		
			_	

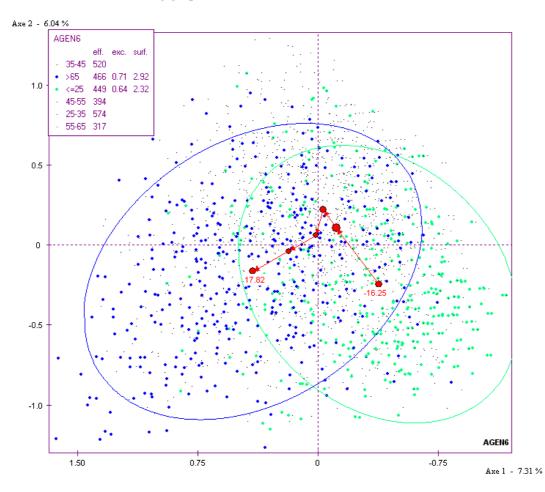
Then put all the points as "ghosts" and click on Total unselction , you get the following graph:



To draw only the ellipses for the young and the old, put the other categories as <u>ghosts</u> (Click on the category 35-45 and then on <u>normal <---> fantôme</u>, etc.



The result is the following graph:



With a click on the right button of the mouse you can modify "by hand" the ellipses, the modality mean points of the categories, or the labels in the graph.

# 4. STORING PRINCIPAL COORDINATES AND PARTITIONS

To recover the principal axes [factorial axes] or the partitions in a SPAD Database, insert the method <u>Storing factorial axes and partitions</u> into the desired location in your chain, then parameterize it and finally execute it.

lect method	
Group of methods :	
Storing, Export	
Descriptive Statistics Factorial Analysis Cluster Analysis Storing, Export Segmentation Decision - Models Multiple tables Amado Textual analyses	<u>O</u> K Cancel <u>H</u> elp
Method : Storing of factorial axes and partitions Data Export Factorial Coordinates Export	
Storing of factorial axes and partitions	

You have two tabs to choose from:

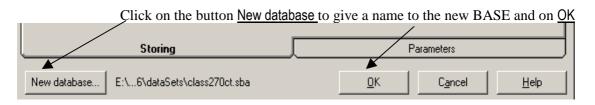
1. The principal coordinates (factorial coordinates) and the partitions to store.

9	TORING OF FA	ACTORIAL COORDINATES AND	PAR
	Storing	Partitions 💌	
	Available pa	Partitions Factorial axes	

2. The parameter settings for archiving.

	3.	
51	TORING OF FACTORIAL COORDINATES AND PARTITIONS	
10		
	Printout parameters	
		Preferences
	New dictionary printout O No O Short O Long	
		(
		Default
	Working parameters	Save
	Continuous variables missing data 💿 Without changing 💽 O Other 1	

The storage parameters have default values.



#### Then run the method.

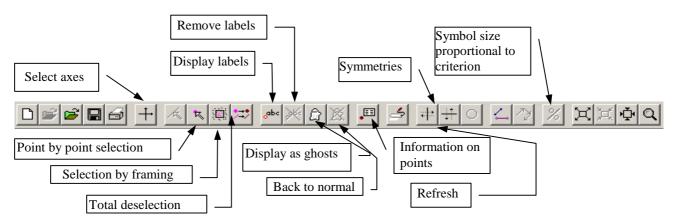
You obtain the following pictogram. In the text file, you can see the name of the variables in the new file and the name of this new file.



# **5. APPENDIX**

## 5.1. Generalities on the graph editor

• The toolbar of the graph editor

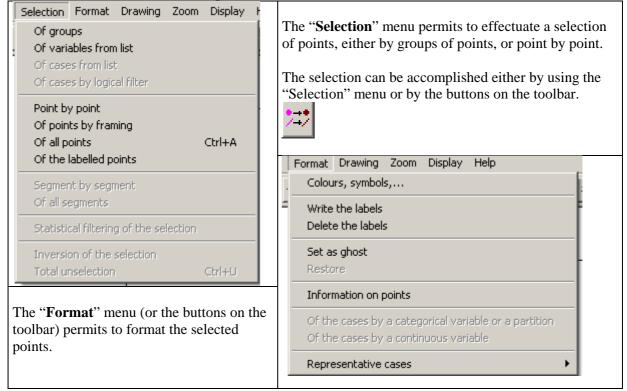


The initial pre-selection for a new graph is important, it is necessary to pre-select the variables and the individuals you think you will be interested in analyzing in the graph.

#### • The fundamental rule for formating a graph

#### Selection /Format (action) / Unselection

One starts with selecting a point or a group of points, then one formats them, and finally one makes a total unselection, and the chain of procedures can continue.



Because certain operations, for example the replacement of labels, are resource demanding, the graph can be imperfect with double labels or blank spots. It is thus advisable to <u>Refresh</u> the graph. Either

click on the icon and the toolbar or use the menu: Drawing/Refresh.

#### 5.2. Interface SPAD/SPSS using SPAD editor

This option allows you to import a file in SPSS format and transform it into a SPAD database, or conversely, to export a SPAD database towards a file in SPSS format.

#### Identifier of observations

In SPSS, there is no specific variable identifying the observations (cases) as such, only an internal variable named \$casenum.

In SPAD, there is a specific variable identifying the observations (cases). In order to allow SPAD to identify the cases of an SPSS database (\*.sav), it is necessary to create the variable id\_spadn, which is a chain of characters recognized by SPAD as the variable defining the cases.

Before importing an SPSS database, create the variable id\_spadn with the help of the following SPSS syntax:

STRING id\_spadn (A4). COMPUTE id\_spadn = STRING(\$casenum,F4) . VARIABLE LABELS id\_spadn 'Identifier SPAD' . EXECUTE . *Management of variables types.* In SPAD, there are two types of variables, <u>NOMINALS</u> (categorical variables) and <u>CONTINOUS</u> (numerical variables).<sup>10</sup> When editing an SPSS database in the data editor of SPAD, the program automatically determines the type of variable according to the following rules:

All numerical variables in SPSS with a range of values between 1 and 300 are automatically ascribed as "nominal variables".

All other numeric variables in SPSS are automatically typed as "continuous variables".

Warning: All SPSS variables with at least one "0" is automatically typed as "continuous variables".

*Management of missing values.* The missing values (by default or specified) in the SPSS database are in the SPAD data editor represented by an empty cell. The program uses respectively the value "0" for categorical variables and 999999 for numerical variables.

*Management of labels and variable names.* For the variables that are automatically typed as "<u>nominal</u> <u>variables</u>", the labels associated with the values (in the SPSS database) are copied into SPAD in the same order. If there are no labels for certain values between the minimum and the maximum value, the program automatically labels the value 'category n° x'. The labels are visible in the window <u>values</u>.

*Warning*: To avoid having a long list of "empty" labels of the type 'category n° x', make sure to code the initial data without gaps. Do not for instance code the variable sex as following, 1 'men', 2 'women' 99 'NR', but 1 'men', 2 'women' 3 'NR'. In SPSS it is possible to "pack" the values by using the <u>automatic recode</u> function.

In the Variable window, the short identifier of a variable is constituted by the first 4 characters of the name of the variable in the SPSS database. The variable label of SPSS is also imported, which makes it easy to identify the variables. Also the initial order of variables in the SPSS database is preserved.

*Management of string variables.* . However, the string variables in the SPSS database can't be imported directly into SPADIt is thus necessary to first transform the string variables into numeric variables in SPSS and then to import them as nominal variables. This can be done by using the automatic recode function (menu Transformation) in SPSS. Use the following syntax to automatically recode, for example, the variable "S" to the new variable "SUJ":

#### AUTORECODE

VARIABLES=S /INTO suj /PRINT.

This function creates a numerical variable which starts at 1 and whose values are arranged in alphabetical order based on the original string variable.

<sup>&</sup>lt;sup>10</sup> There is also a third type, TEXT (text or string variables), but it is not relevant in the context of importing SPSS databases, see below.

• To Import SPSS databases: \*.sav (SPSS) → \*.sba (SPAD)

1. In the main window of SPAD, chose Dataset and New database to enter the SPAD database editor.

SPAD					
Dataset	Chain	Tools	Options	Help	
New d	ataset				
Edit da Edit lai	ataset bel	Ctr	I+T		
Import ►					
Export •					
Delete dataset					
Exit					

2. In the menu File in the data editor, click on Open

📺 Data Editor					
File View ?					
New	Ctrl+N				
Open	Ctrl+O				
Recent file					
Preferences Continuous missing data					
Exit					

3. Select Files of type: SPSS file in the list of file types.

Ouvrir		? ×
<u>R</u> egarder dans :	🔁 BasesSPSS 💽 🗲 🖻 📸 📰 🗸	
Nom <u>d</u> u fichier :	<u></u>	<u>D</u> uvrir
Fichiers de <u>t</u> ype :	Base file (*.sba)	nnuler
	Base file (*.sba) SPSS file (*.sav)	
	SPSS tile (* sav) Tous les fichiers (*.*)	

4. Then search for and select the SPSS database (\*.sav) you want to import (here Culture.sav)

Ouvrir	? >	1
<u>R</u> egarder dans	: 🔁 BasesSPSS 💽 🗢 🔁 📸 🖬 🕶	
culture.sav	,	
	Type : Document de données SPSS Date de modification : 21/08/2006 18:39 Taille : 55.9 Ko	
Nom <u>d</u> u fichier :	culture.sav Quvrir	
Fichiers de <u>t</u> ype	: SPSS file (*.sav) Annuler	

The SPSS database is automatically opened in the SPAD data editor.

Dal	ta Editor - 🛛	culture.sav	_ 🗆 ×
jie <u>p</u>	<u>i</u> dit ⊻iew	Window 2	
D (	🎽 🖬   🖇	X 🗈 🛱 🖬 🖬 🖬 🛤 🔤 🔤 🔤 🖬 🖬 🖬 🖓	
Var CU	lture.sav :	: Variables	
	Ident	Label T Min Max 🔺 Ident Label	
	lden	Identifiant T T	
	Libl	Libellé T T	
1	Q1	do you prefer leisure activities you ce N 1	
2	Q2	Would you say that during your free ti N 1	
3	Q3	If you had more time, 1st choice woul N 1	
4	Q4	When you go out in the evening, do j N 1	
5	Q5	Time watching TV (hours by week) N 1	
•			
		24 culture.sav : Values	
		Iden Libl Q1 Q2 Q3 Q4 Q5 Q6 QS1 QS2 QS3 QS4 Z	
		2 2 4 1 4 1 3 3 1 3 6 10 1	
		3 3 4 1 7 2 4 2 1 7 1 8 1	
		4 4 1 1 2 3 1 1 6 1 11 1	
		5 5 1 1 5 5 5 2 2 1 4 9 1	
		<u>6</u> <u>6</u> <u>3</u> <u>2</u> <u>3</u> <u>5</u> <u>5</u> <u>1</u> <u>2</u> <u>2</u> <u>3</u> <u>9</u> <u>1</u>	
	help, press F		
goe	10,p10351		

To create an SPAD file, chose Save as and select the SPAD file type (.sba) in the list of file types. Name the database (here Culture).

*Warning*: Sometimes in the box for the name, the file extension \*.sav appears. It is then necessary to erase this and change it to \*.sba, otherwise the initial SPSS database will be overwritten with the SPAD file!

🚺 Data Editor - culture.sav		Enregistrer sous		
File Edit View Window ? New	Ctrl+N	Enregistrer <u>d</u> ans :	🛅 BasesSPSS	▼ ⇐ 🗈 💣 🎟 -
New				
Open	Ctrl+O			
Close				
Save	Ctrl+S			
Save as				
1 E:\data\BasesSPSS\culture.sav				
Preferences				
Exit				
		Nom <u>d</u> u fichier :	culture	<u>1</u>
		<u>T</u> ype :	Base file (*.sba)	<b>~</b>

The file culture.sba which now is created is the database for SPAD.

*Important remark:* The database is not modifiable in SPAD. In case of problem with the import, make the corrections in SPSS and redo the import SAV  $\rightarrow$  SBA.

• To Export SBA databases: \*.sba (SPAD) → \*.sav (SPSS)

To export a SPAD database to a file in the SPSS format, select the option **Edit Dataset** in the **Dataset** menu, then **File-Open**.

- ① Select the base to export and it will be opened in the Database editor.
- ② Select the option **Save** as in the **File** menu.
- ③ Select the SPSS file type (\*.sav) and enter the name of the file to create (by selecting the extension .sav), then click on Save.

#### • Rules for converting SPAD to SPSS

SPSS identifies the variables by their name. The conversion of SPAD to SPSS follows the rules described below.

• Categorical variables.

The <u>categorical variables</u> are copied into SPAD with their labels and values identifier, with the value 0 for the missing data.

• Numerical variables.

The <u>continuous variables</u> are copied with their labels, and the value 9999999 for the missing data. • The variable names are the short identifiers. The *labels* (<u>view variables</u>) are copied to the 'variable tag' in SPSS.

• Identification of cases.

The content of the label column of cases is sent to SPSS in the form of a character string with the name id\_spadn.

• Duplicates of variable names.

In the SPAD database editor, the short identifier is the string containing the first four characters of the label. In the case of these duplicates, the name of the variable in SPSS is created by attaching to the short identifier an index position of the variable in the SPAD file.Variable names must begin with a letter.