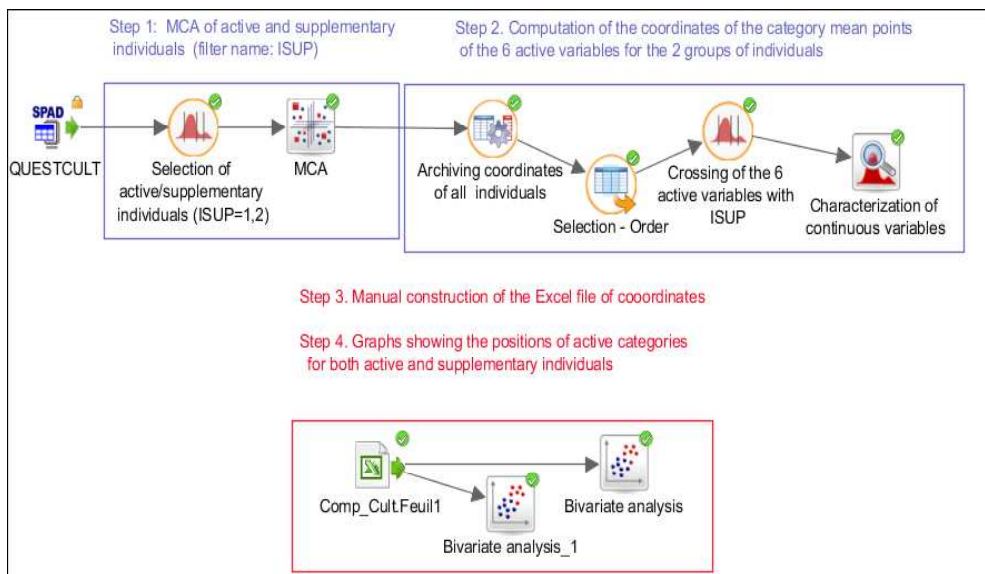


**Question 1:** *In order to do a comparative analysis I wish to project the supplementary dataset into the analysis. I would like to show the positions of categories associated with supplementary individuals alongside the active categories of active individuals.*

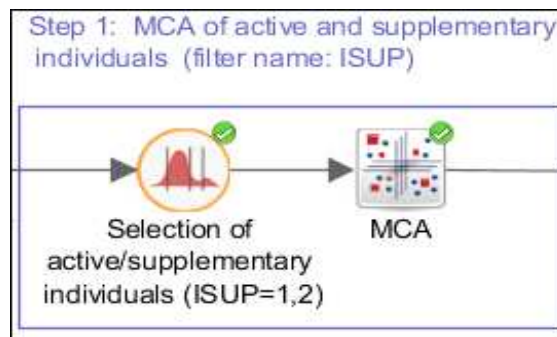
**Answer:** *The solution is to build the mean points of subclouds associated with each active category in the cloud of individuals. This must be done on the one hand for active individuals and on the other hand for supplementary individuals. Then estimate the deviation between the two corresponding categories.*

To obtain the solution with SPAD on your dataset, reproduce the process of the “archived project” called “MCA\_Q1\_Jan2010.spad” (to be downloaded). The diagram associated with the project is the following one.



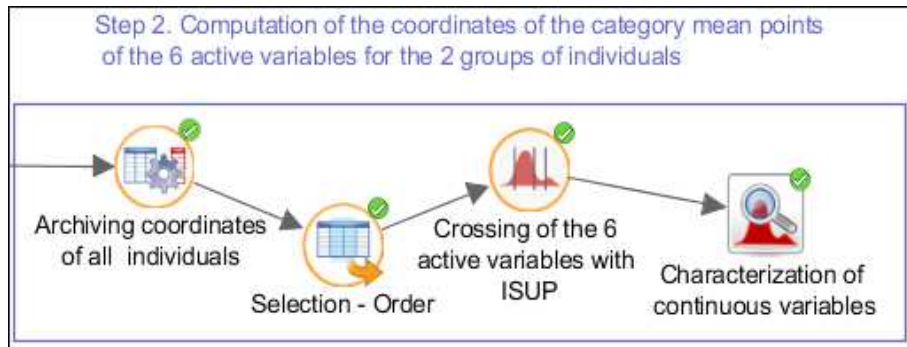
As an example, we used the data of the Culture Example presented in the book “Geometric Data Analysis”<sup>1</sup> (chapitre 6, page 221).

- The **first step** consists of 2 points:



<sup>1</sup> Brigitte Le Roux & Henry Rouanet, *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*, Kluwer, Dordrecht, 2004.

1. Construct a filter (called ISUP) to select the active individuals (those who are over 25) and the supplementary individuals (the others).
  2. Perform MCA with the over 25 as active individuals and the under 25 as supplementary ones.
- The second step consists in 4 points:



1. Archive the coordinates of individuals (the last two columns of the data table contain the coordinates of *both* active and supplementary individuals).
2. Put into order the variables with the filter ISUP in second position<sup>2</sup>.
3. Cross the 6 active variables with the filter ISUP (Data management>Columns – Variables>Bining – Grouping – Crossing). Hence 2x6 new variables.
4. Calculate the coordinates of the category mean points of the categories of the 2x6 variables (Descriptive statistics>Bivariate analysis : the continuous variables to characterize are the principal axes, the Characterizing categorical variables are the 2x6 new variables).

In folder 1, sheet Desco-A, you obtain the coordinates (highlighted in yellow) of the category mean points for active individuals and the coordinates (highlighted in blue) of the supplementary individuals who have chosen those categories as well as their weights.

Overall variance		0.295		
Variable label	Category	Mean	Weight	Between var variance and C
ISUP_do yo		-0.06	3002.00	
	<=25_LeisureAlone	0.01	40.00	
	<=25_LeisureFamily	-0.03	41.00	
	<=25_LeisureFriends	-0.38	438.00	
	<=25_LeisureNA	-0.53	1.00	
	<=25_LeisurePartner	-0.31	65.00	
	>25_LeisureAlone	0.35	436.00	
	>25_LeisureFamily	0.03	708.00	
	>25_LeisureFriends	-0.18	774.00	
	>25_LeisureNA	0.50	20.00	

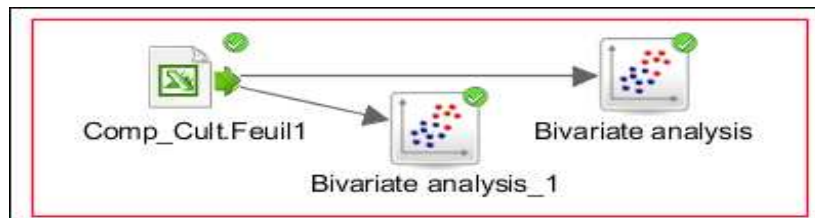
Desco-1 / **Desco-A** / Desco-2 / Desco-3 / Desco-4

<sup>2</sup> This point is not essential but is aimed at making the manual construction of the Excle file easier (step 4).

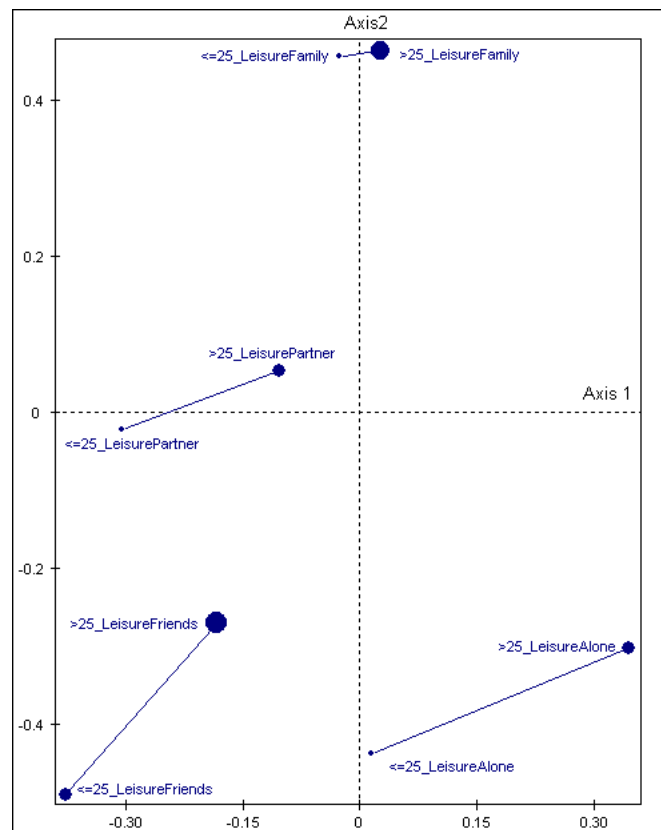
- The **third step** consists in constructing manually the Excel file named Comp\_Cult.xls from the sheets Desc-A of folders 1 and 2. This file consists of the labels of categories (after crossing); the coordinates of the categorie mean points on axis 1 and on axis 2 as well as their weights as follows:

Comp_Cult.xls				
	A	B	C	D
1	Category	Axis 1	Axis2	Weight
2	<=25 LeisureAlone	0.01416	-0.43749	40
3	<=25 LeisureFamily	-0.02672	0.45672	41
4	<=25 LeisureFriends	-0.37739	-0.49116	438
5	<=25 LeisurePartner	-0.30522	-0.02177	65
6	>25 LeisureAlone	0.34554	-0.30242	436
7	>25 LeisureFamily	0.02754	0.46444	708
8	>25 LeisureFriends	-0.18422	-0.27025	774

- The **fourth step**, consists in importing the data from the Excel file (Importing/Exporting data>ImportExcel datasheet), then in carrying out bivariate analyses (Descriptive statistics>Bivariate analysis).



The “Bivariate analysis” permits the construction of the graph of 2x5 categories of the question “leisure” (see the following figure).



The “Bivariate analysis\_1” permits to put on the same graph the questions “Books” and “TV”.

**Question 1bis.** *How to compute with SPAD the positions of supplementary variables for supplementary individuals.*

The solution is given by this diagram called question 2. We apply the solution on the Culture example for the supplementary individuals who are under 25.

