

# 1. Topic

## Installation and use of the "sipina.xla" (Sipina version 3.12) add-in for Excel 2007.

SIPINA is a Data Mining Software which implements various supervised learning paradigms. This is an old tool but it is still used because this is the only free tool which provides fully functional interactive decision tree capabilities.

This tutorial briefly describes the installation and the use of the add-in "sipina.xla" into Excel 2007. The approach is easily generalized to Excel 2010. A similar document exists for Tanagra<sup>1</sup>. It seemed to me nevertheless necessary to clarify the procedure, especially because several users have made the request. Other tutorials exist for earlier versions of Excel (1997-2003)<sup>2</sup> and for Calc (Libre Office and Open Office)<sup>3</sup>. A new tutorial will come soon. It shows that the add-in performs finely also under Excel 2016.

## 2. Installation of Sipina

The first step is to install Sipina on our computer if it is not already made. The Sipina packages comprises tools for supervised learning (Sipina), for regression ([REGRESS](#)), and for association rule learning ([ARS](#)). The setup file can be downloaded from the Sipina website : <http://eric.univ-lyon2.fr/~ricco/sipina.html>

The screenshot shows a web browser window displaying the Sipina website. The page has a dark green background and contains several sections:

- Data Mining Links:** French Data Mining Portal, Tanagra Software, Ricco's web site.
- SIPINA DOWNLOAD:** Includes a link to the French website.
- Software:** Most recent Sipina Research version - 32 bits. Implements several supervised learning methods (decision tree, neural network, linear discriminant analysis,...), model assessments (cross-validation, bootstrap,...) and association rule algorithm. A red arrow points to the 'Sipina Research' file link.
- Documentation:**
  - SIPINA Add-in for EXCEL spreadsheet:** (In english) An add-in for EXCEL(c) is incorporated in the SIPINA distribution. This add-in (SIPINA.XLA) enables to start a classification tree analysis, and more generally a data mining process, from your spreadsheet. This classification tree add-in appends a new menu in your spreadsheet. You select the cells range, activate the right menu: SIPINA is started and the selected dataset is automatically loaded.
  - Building decision tree interactively for the analysis of high blood pressure with SIPINA.** Tutorial
  - Using predefined learning (training) and test set for classifier performance evaluation.** Tutorial
  - Definition of misclassification costs and the utilization of cost-sensitive**

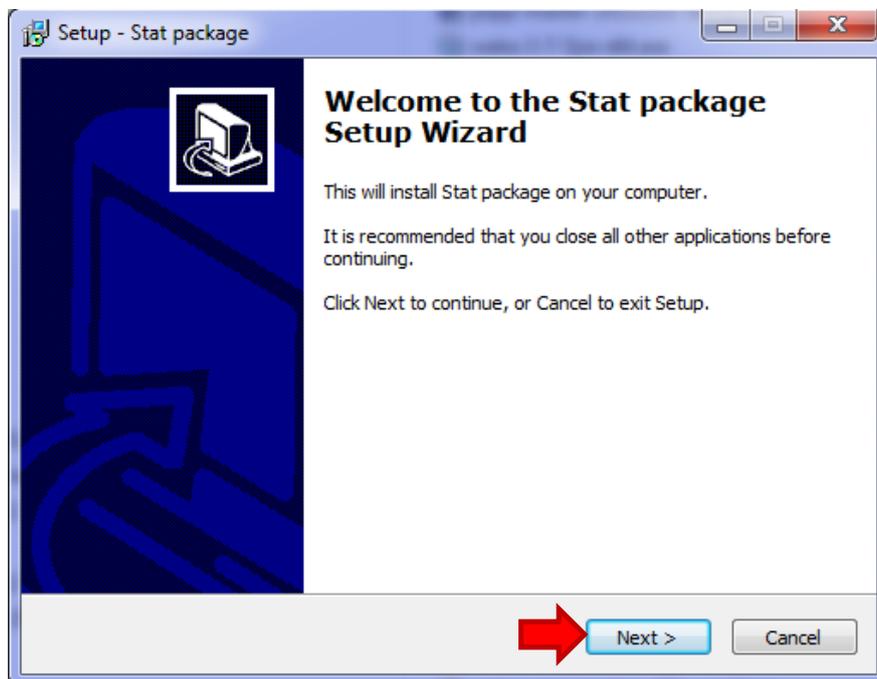
On the left side, there is a sidebar with a 'Research' logo and links for 'Sipina Features' (Overview, Main Features, Learning Algorithms) and 'Sipina availability' (Download Sipina). A red arrow points to the 'Download Sipina' link. At the bottom left, there is a 'Version française' link with a French flag icon.

<sup>1</sup> <http://data-mining-tutorials.blogspot.fr/2010/08/tanagra-add-in-for-office-2007-and.html> (august 2010)

<sup>2</sup> <http://data-mining-tutorials.blogspot.fr/2010/08/sipina-add-in-for-excel.html> (august 2010)

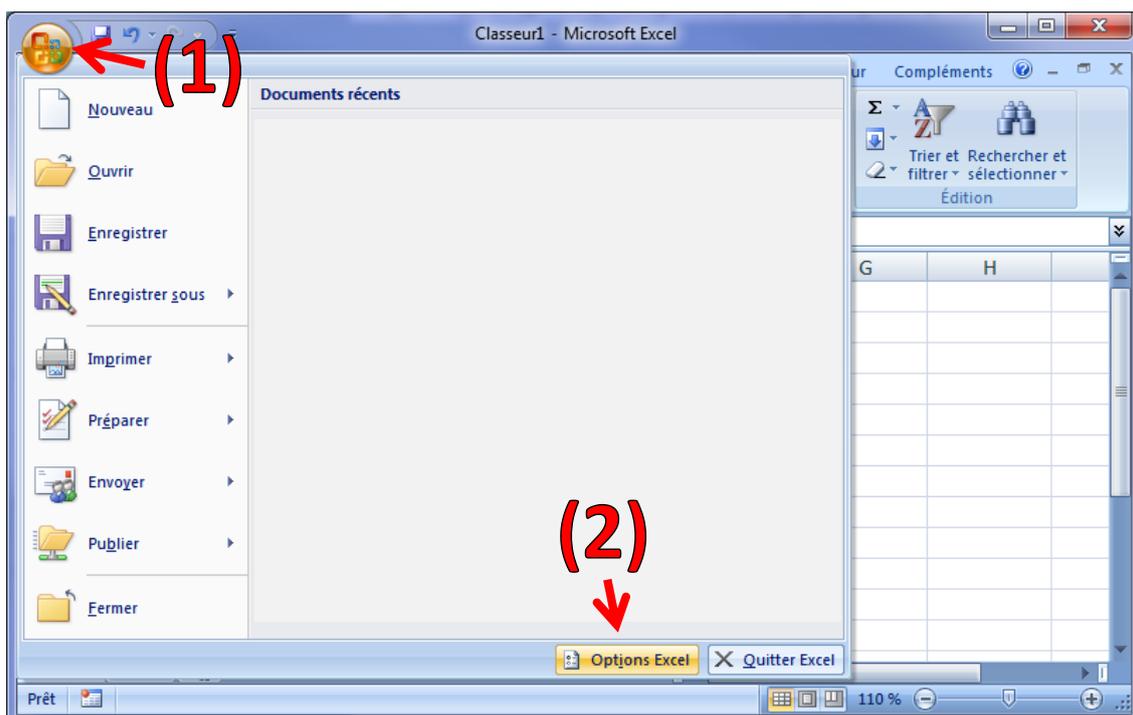
<sup>3</sup> <http://data-mining-tutorials.blogspot.fr/2012/03/sipina-add-on-for-oocalc.html> (march 2012)

After we have downloaded the setup file, we start the installation. This is a standardized procedure. You can simply click NEXT throughout the process.

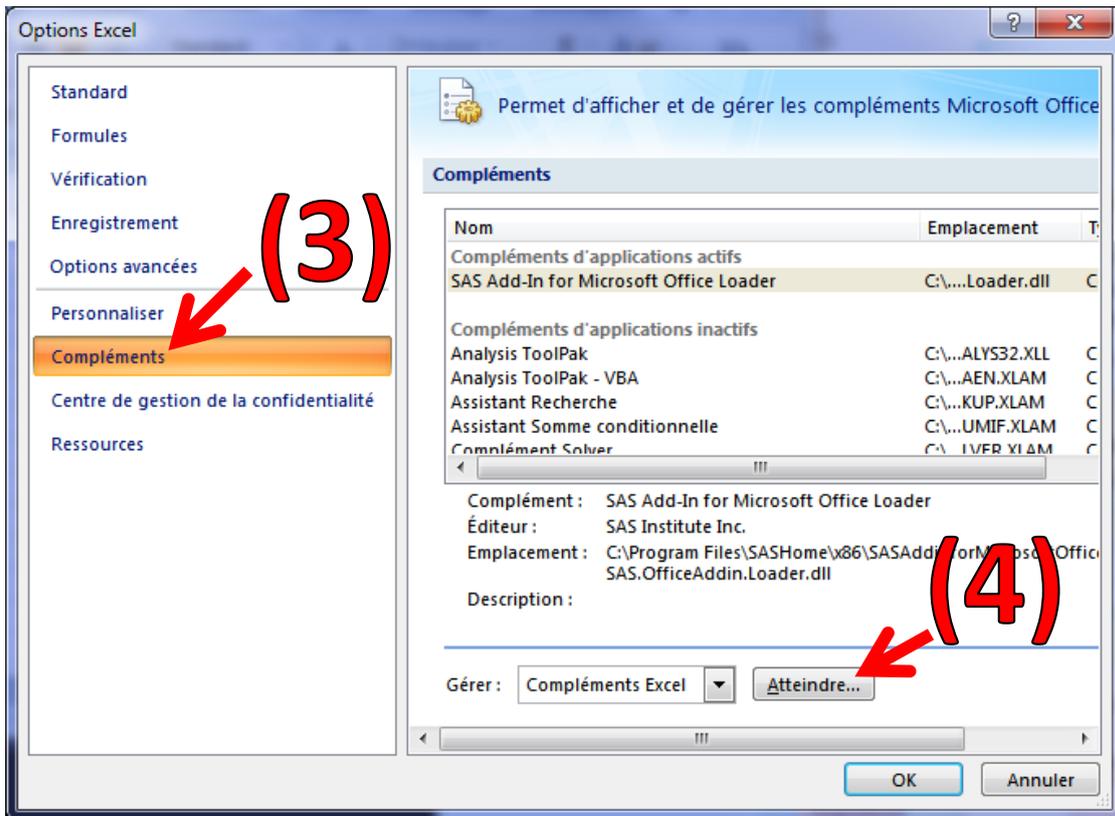


### 3. Installation of the add-in in Excel 2007

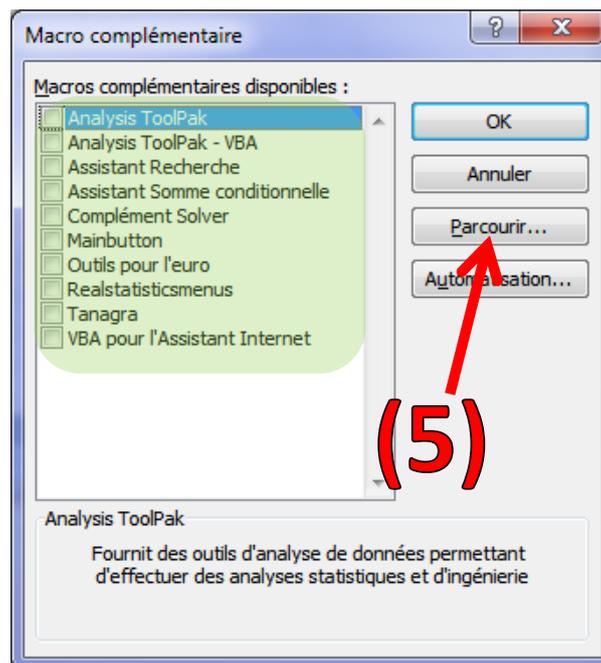
After we launch Excel, we click on the Office button (1), then we click on "Excel options" (2). The menus are in French into the screenshots, but I think you can easily perform the same process whatever the language of your version.



Into the setting dialog, we select "add-ins" option (3), then we click on the "Go" button (4).

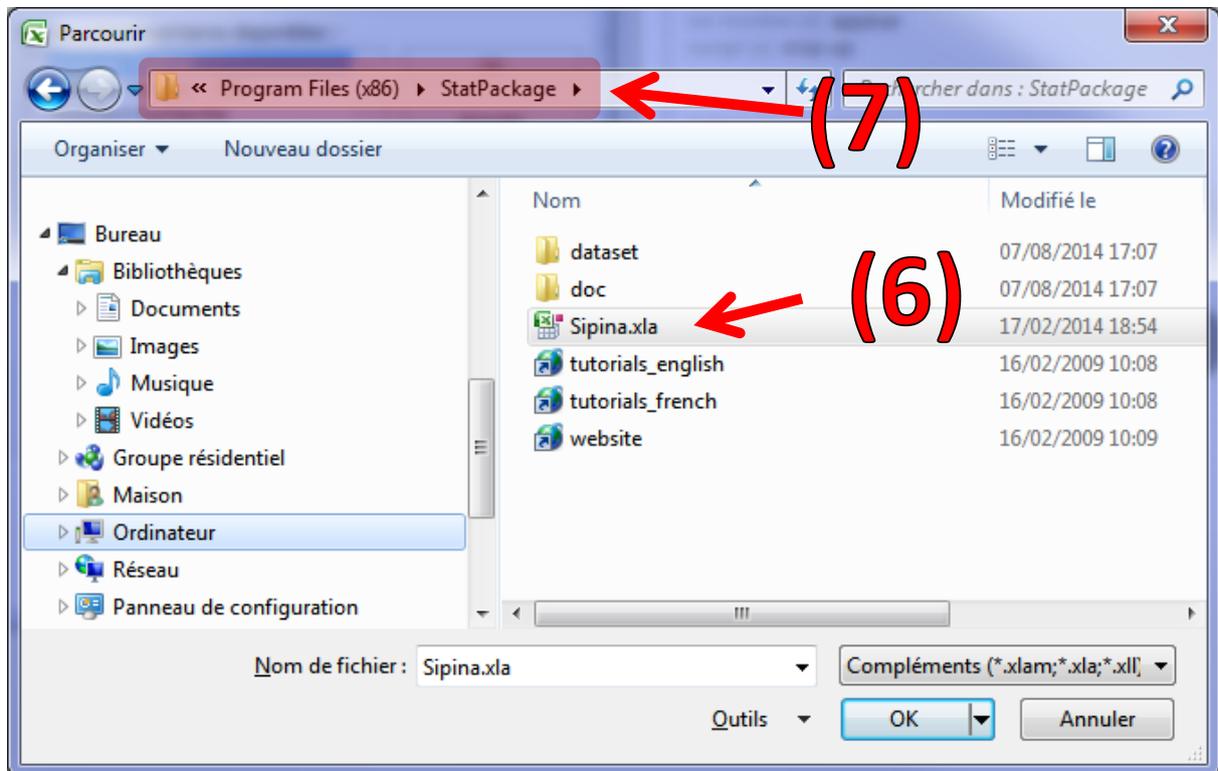


The list of available add-ins appears in the add-in dialog box. For a first installation, the checkbox Sipina is not visible. We click on the "browse" button in order to select the "sipina.xla" file from our computer.

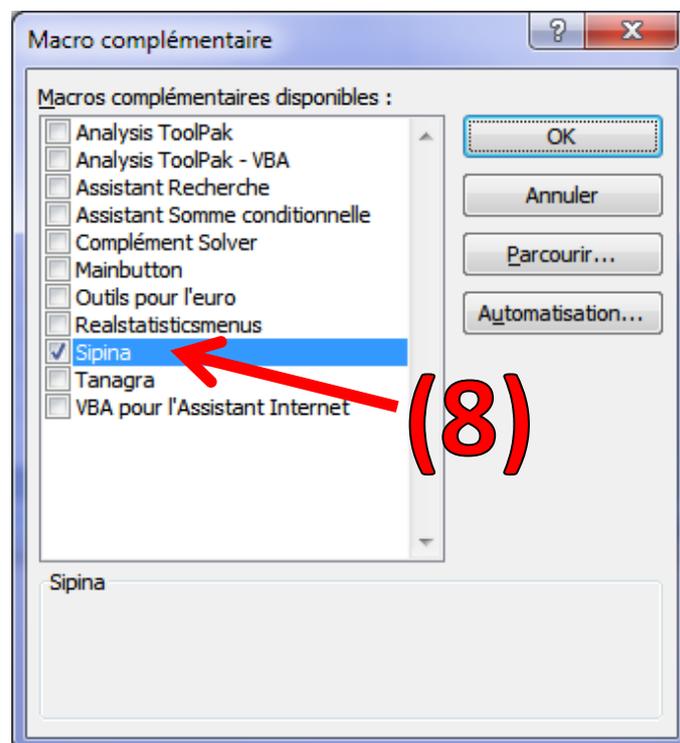


The add-in « **Sipina.xla** » (6) is accessible in the folder:

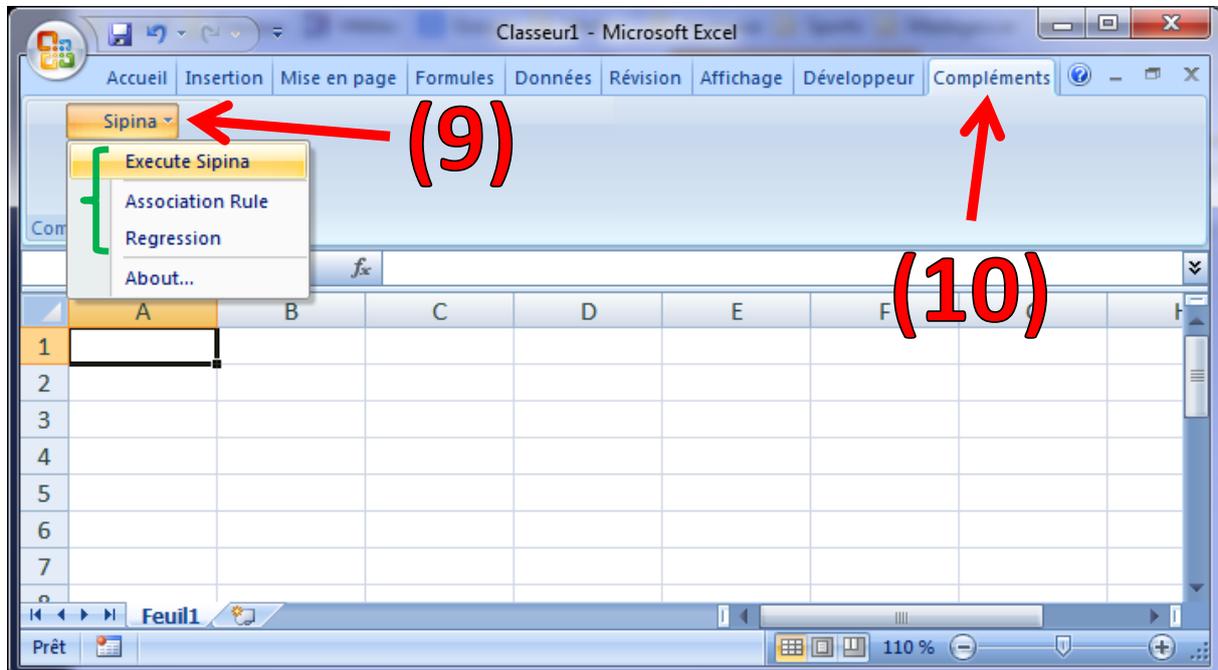
- « c:\Program Files\StatPackage » for 32-bit Windows;
- « c:\Program Files (x86)\StatPackage » for 64-bit Windows (7).



The add-in is installed, it is visible and activated in the add-in dialog box (8).

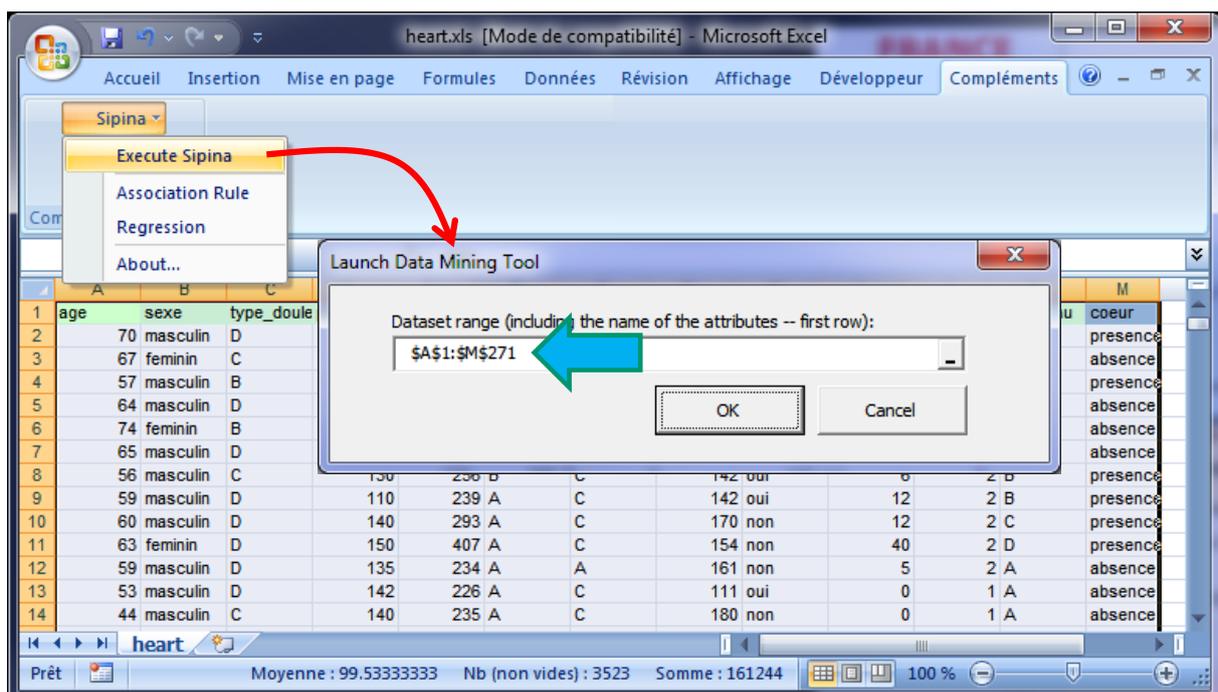


We click on the OK button. The menu SIPINA (9) is now available in the "add-ins" tab (10). The three modules are visible: *Sipina* intended to decision tree learning; *Association rule* for association rule learning; and *Regression* for regression analysis.



#### 4. Use of the add-in

The use of the add-in is described on many occasions in the tutorials about SIPINA. We go to the basics in this section.



We load the "heart.xls" data file. We select the data range and we click on the SIPINA / EXECUTE SIPINA menu. A dialog box allows to check the coordinates of the data range (\$A\$1:\$M\$271). The first row must correspond to the names of the variables.

Sipina is launched. The data are automatically loaded. We have 13 variables and 270 instances. We are ready for a new analysis.

	age	sexe	type_douleur	pression	cholester	sucre	electro	taux_max	angine	depression	pic	vaisseau	coeur
1	70.00	masculin	D	130.00	322.00	A	C	109.00	non	24.00	2.00	D	presence
2	67.00	feminin	C	115.00	564.00	A	C	160.00	non	16.00	2.00	A	absence
3	57.00	masculin	B	124.00	261.00	A	A	141.00	non	3.00	1.00	A	presence
4	64.00	masculin	D	128.00	263.00	A	A	105.00	oui	2.00	2.00	B	absence
5	74.00	feminin	B	120.00	269.00	A	C	121.00	oui	2.00	1.00	B	absence
6	65.00	masculin	D	120.00	177.00	A	A	140.00	non	4.00	1.00	A	absence
7	56.00	masculin	C	130.00	256.00	B	C	142.00	oui	6.00	2.00	B	presence
8	59.00	masculin	D	110.00	239.00	A	C	142.00	oui	12.00	2.00	B	presence
9	60.00	masculin	D	140.00	293.00	A	C	170.00	non	12.00	2.00	C	presence
10	63.00	feminin	D	150.00	407.00	A	C	154.00	non	40.00	2.00	D	presence
11	59.00	masculin	D	135.00	234.00	A	A	161.00	non	5.00	2.00	A	absence
12	53.00	masculin	D	142.00	226.00	A	C	111.00	oui	0.00	1.00	A	absence
13	44.00	masculin	C	140.00	235.00	A	C	180.00	non	0.00	1.00	A	absence
14	61.00	masculin	A	134.00	234.00	A	A	145.00	non	26.00	2.00	C	presence
15	57.00	feminin	D	128.00	303.00	A	C	159.00	non	0.00	1.00	B	absence
16	71.00	feminin	D	112.00	149.00	A	A	125.00	non	16.00	2.00	A	absence
17	46.00	masculin	D	140.00	311.00	A	A	120.00	oui	18.00	2.00	C	presence
18	53.00	masculin	D	140.00	203.00	B	C	155.00	d	31.00	3.00	A	presence
19	64.00	masculin	A	110.00	211.00	A	C	144.00	o	18.00	2.00	A	absence

## 5. Conclusion

In this tutorial, we propose a brief description of the installation and the use of the add-in "sipina.xla". For computer scientists, the source code is available in the Excel's Visual Basic Editor. The insiders will see that the program is relatively simple. I use the Windows clipboard to transfer data. The same principle is used for Tanagra software. This approach remains operational with LibreOffice or Open office Calc under Linux<sup>4</sup>.

<sup>4</sup> <http://data-mining-tutorials.blogspot.fr/2009/04/launching-tanagra-from-oocalc-under.html> for Tanagra.