

1. Objectif

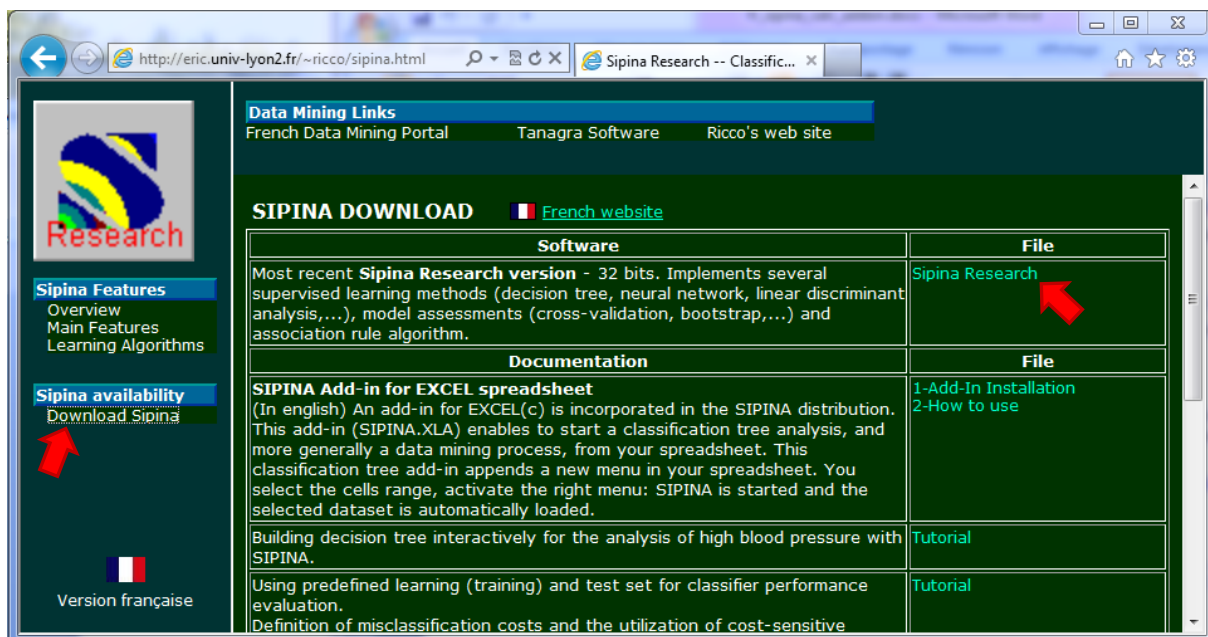
Installation de l'add-in « sipina.xla » (de Sipina version 3.12) pour Excel 2007 et 2010.

Ce tutoriel décrit succinctement l'installation de l'add-in « sipina.xla » dans Excel 2007. La transposition à Excel 2010 est immédiate. Un document similaire existe pour Tanagra¹. Il m'a semblé néanmoins nécessaire de repréciser la démarche, notamment parce que plusieurs internautes m'en ont fait la demande. D'autres didacticiels existent pour les versions antérieures d'Excel (1997 à 2003)² et pour le tableur Calc d'Open Office et de Libre Office³.

2. Installation de Sipina

S'il n'est pas installé sur votre machine, la première étape consiste à télécharger et à installer le « package » SIPINA puisque, outre le logiciel SIPINA proprement dit (spécialisé dans l'apprentissage supervisé, avec une très forte prédilection pour les arbres de décision), il comporte d'autres modules : l'outil ARS (Association Rule Software) pour la construction des règles d'association⁴ ; le logiciel REGRESS pour la régression linéaire multiple⁵.

Le setup du logiciel SIPINA est accessible en ligne <http://eric.univ-lyon2.fr/~ricco/sipina.html>



Après avoir téléchargé le fichier, nous démarrons l'installation. Elle est complètement standardisée. On peut se contenter de cliquer sur NEXT tout au long du processus.

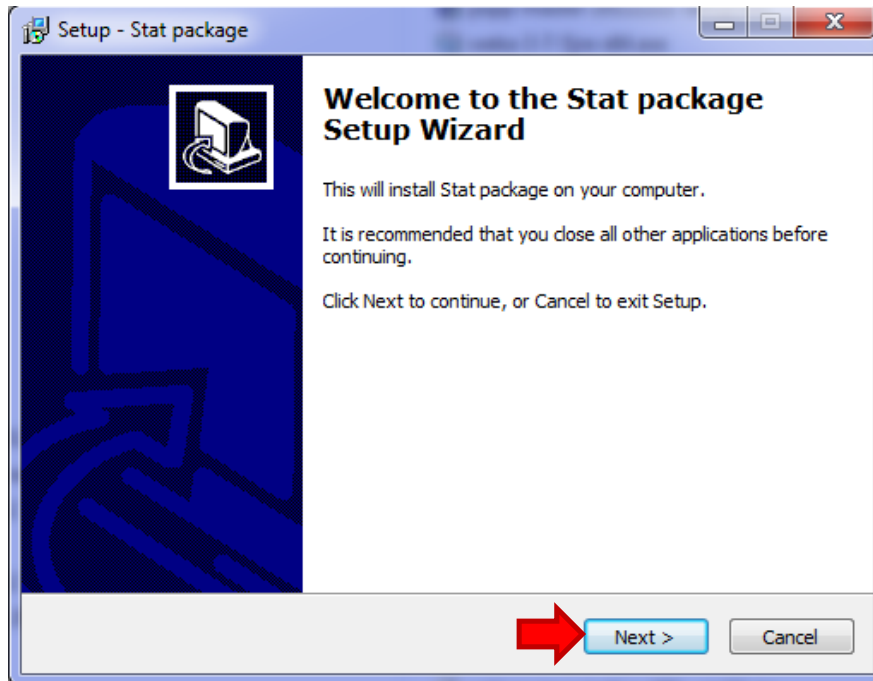
¹ <http://tutoriels-data-mining.blogspot.fr/2010/08/ladd-in-tanagra-pour-excel-2007-et-2010.html>

² <http://tutoriels-data-mining.blogspot.fr/2008/03/connexion-excel-sipina.html>

³ <http://tutoriels-data-mining.blogspot.fr/2012/03/sipina-add-on-pour-oocalc.html>

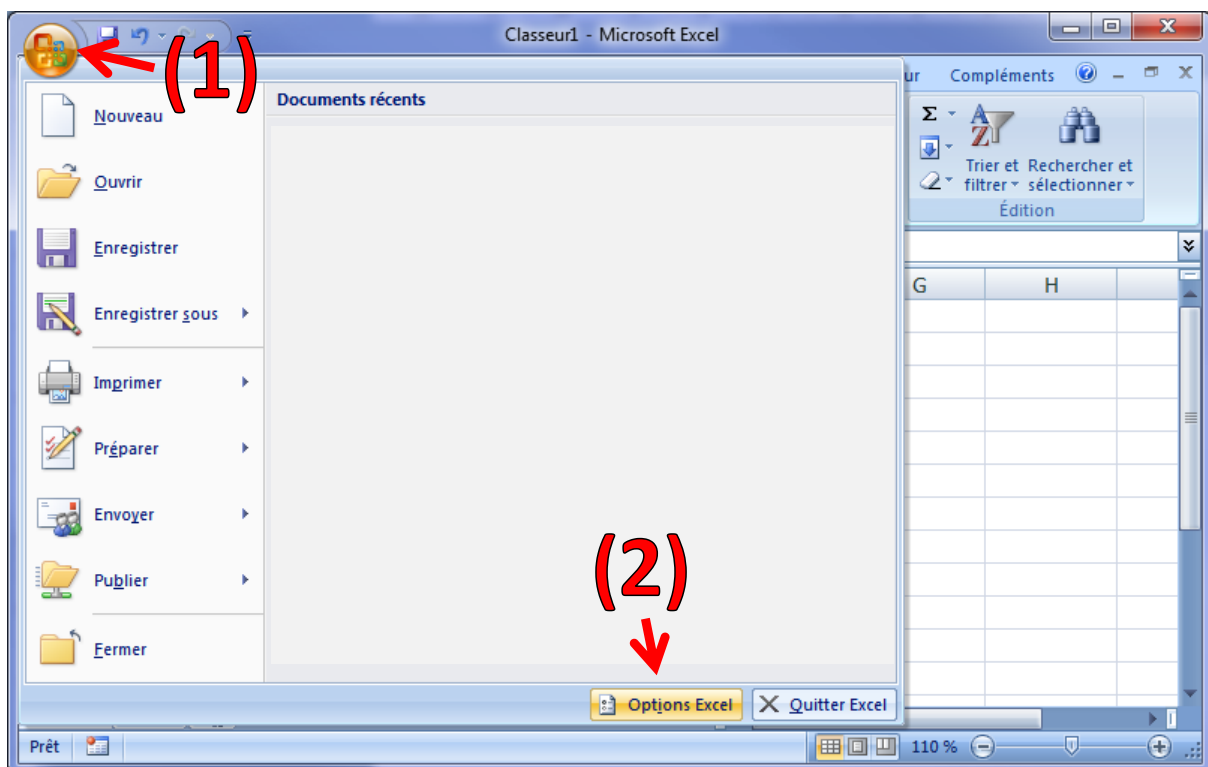
⁴ <http://tutoriels-data-mining.blogspot.fr/2013/04/associations-dans-la-distribution-sipina.html>

⁵ <http://tutoriels-data-mining.blogspot.fr/2011/05/regress-dans-la-distribution-sipina.html>

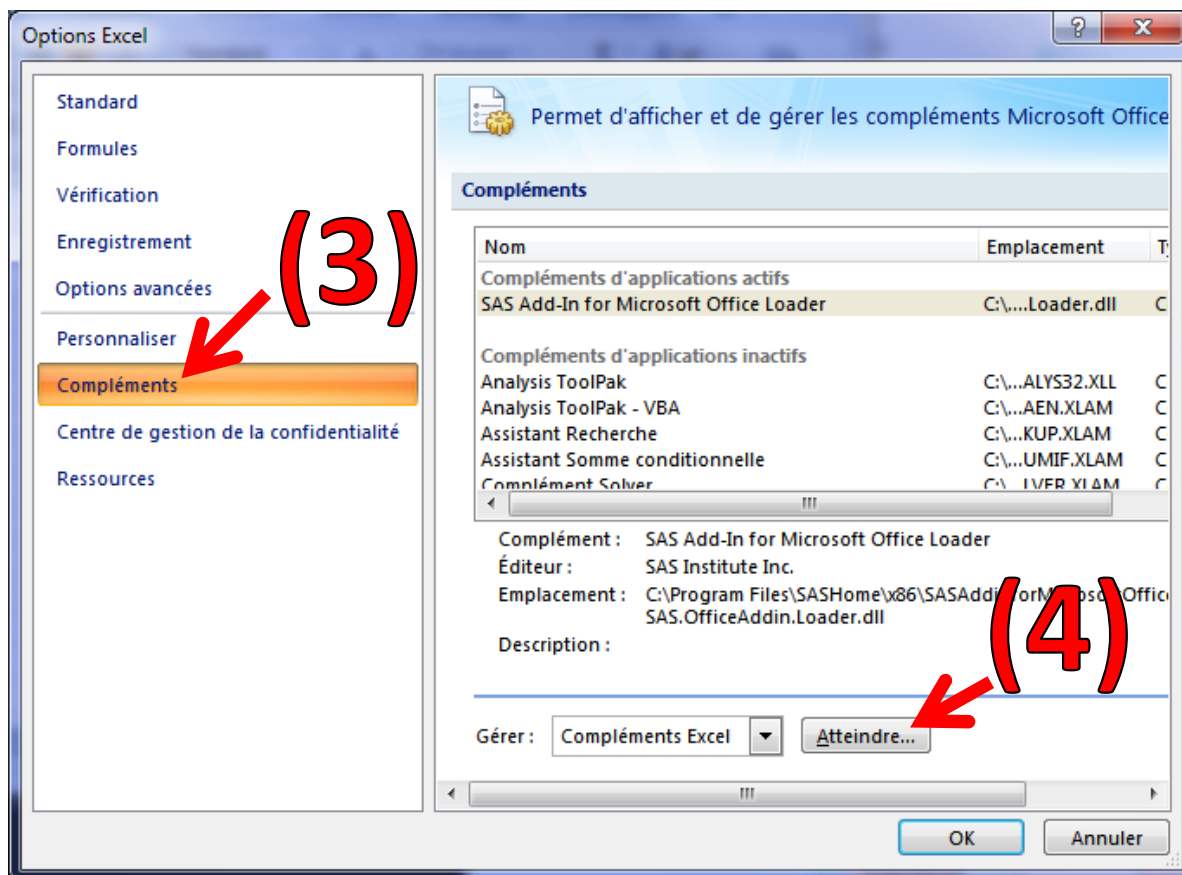


3. Installation de l'add-in dans Excel 2007

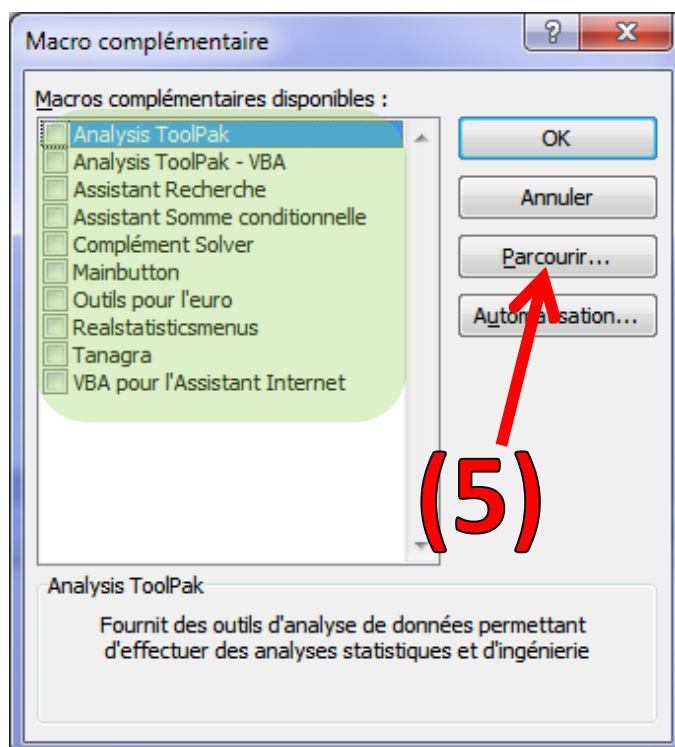
Après avoir démarré Excel, nous actionnons le bouton Microsoft Office (1) et nous cliquons sur « Options Excel » (2).



Dans la boîte de paramétrage, nous sélectionnons l'option « Compléments » (3) et nous cliquons sur le bouton « Atteindre » (4).

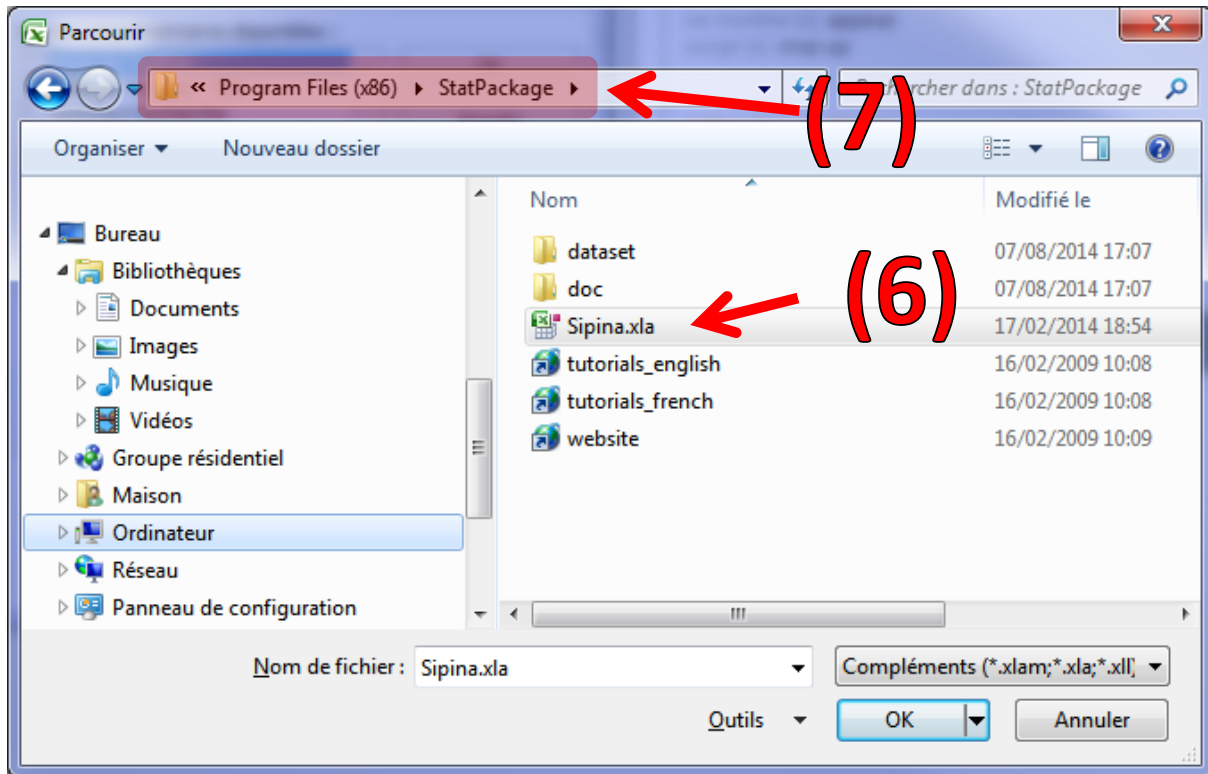


La fenêtre « Macro complémentaire » est visible, avec la liste des add-ins recensés. Pour une première installation, Sipina n'y apparaît pas. Nous cliquons sur le bouton « Parcourir » (5) pour le sélectionner sur le disque de notre machine.

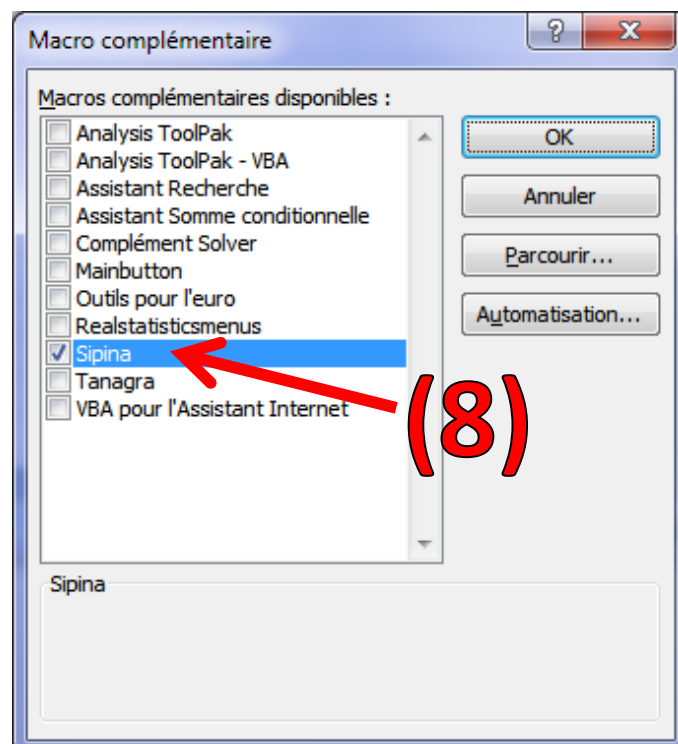


L'add-in « **Sipina.xla** » (6) est accessible dans le répertoire :

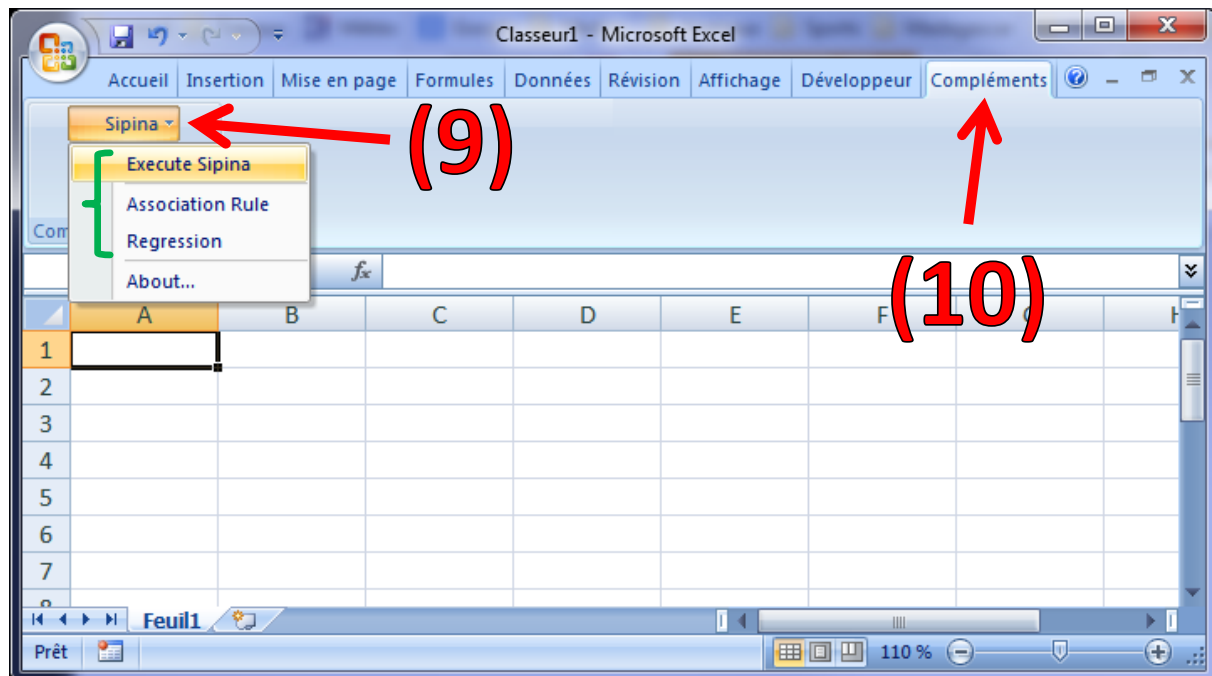
- « **c:\Program Files\StatPackage** » sous **Windows 32 bits** ;
- « **c:\Program Files (x86)\StatPackage** » sous **Windows 64 bits** (c'est le cas pour ma machine) (7).



L'add-in est installé, il est visible et activé dans la fenêtre « Macro complémentaire » (8).

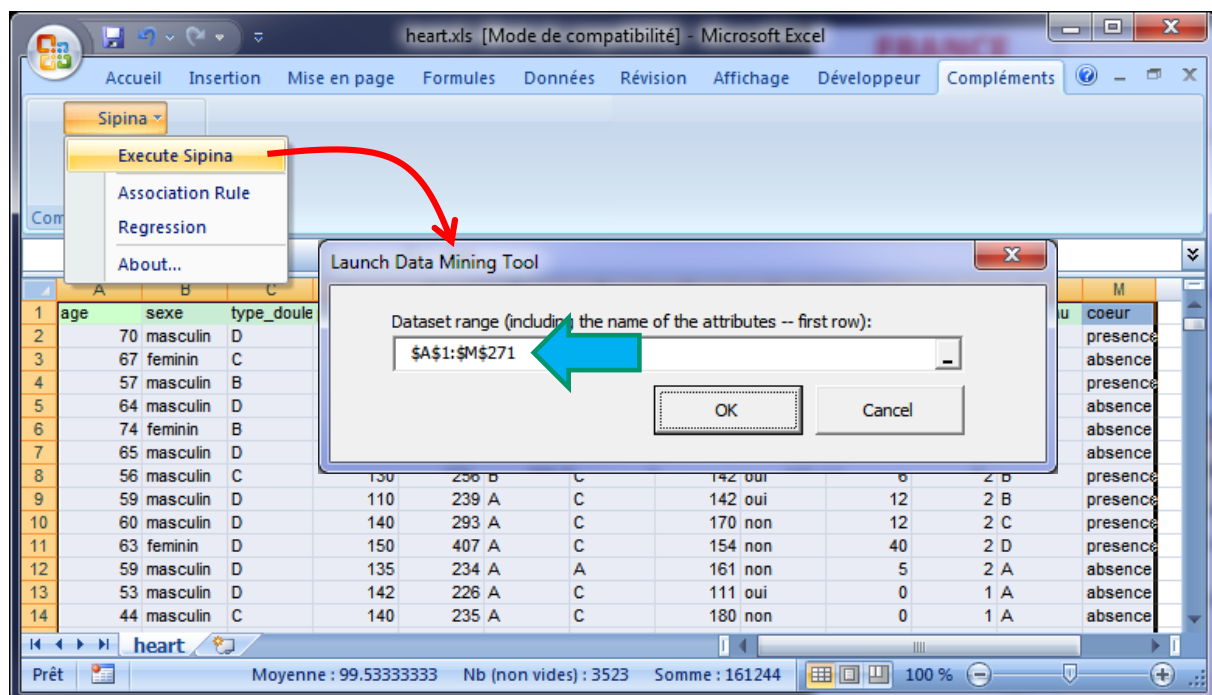


Nous validons. Le menu **SIPINA (9)** est maintenant disponible dans l'onglet **COMPLÉMENTS (10)** d'Excel. Les 3 modules composant la distribution y sont visibles : **Sipina**, spécialisé dans les arbres de décision ; **Association rule**, dédié à l'extraction des règles d'association ; **Regression**, dédié à la régression linéaire multiple.



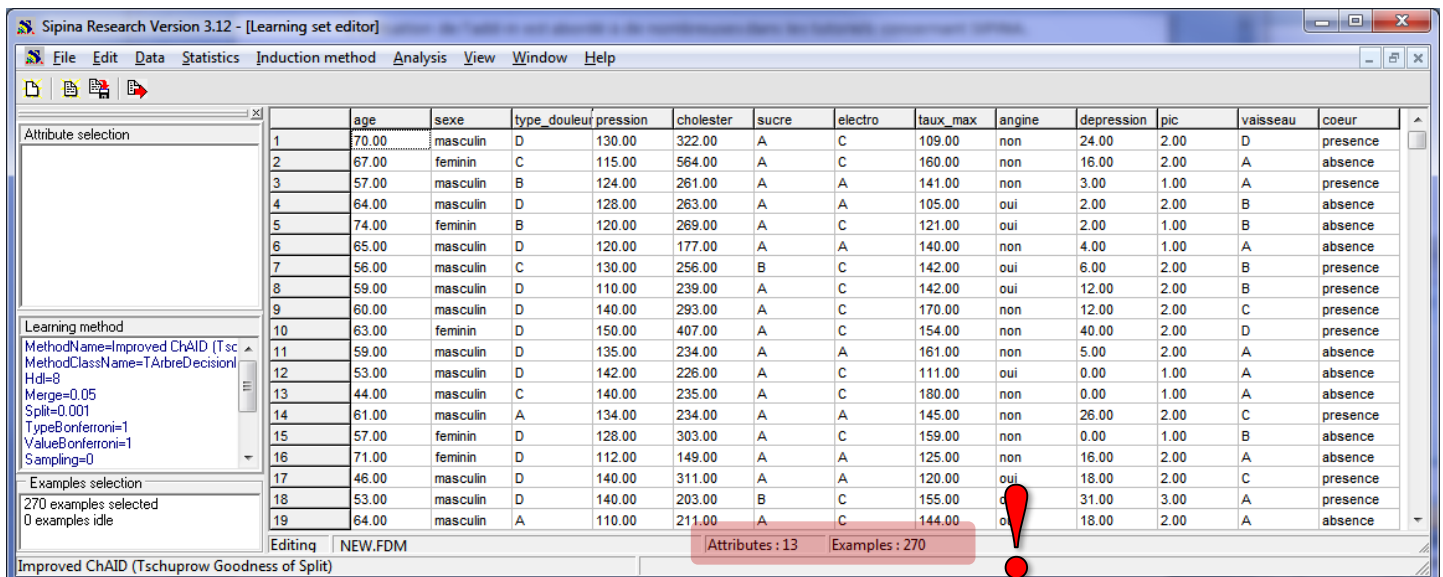
4. Utilisation de l'add-in

L'utilisation de l'add-in est abordé à de nombreuses reprises dans les tutoriels concernant SIPINA. Nous allons à l'essentiel ici.



Nous chargeons le fichier « heart.xls ». Nous sélectionnons la plage de données et nous cliquons sur l'item « Execute Sipina » du menu « Sipina ». Une boîte de dialogue permet de vérifier, et re-spécifier au besoin, les coordonnées de la plage de données (\$A\$1:\$M\$271). La première ligne représente les noms des variables.

SIPINA est démarré. Les données sont automatiquement chargées. Nous avons bien 13 variables et 270 observations. Nous sommes prêts pour une nouvelle analyse.



| | age | sexe | type_douleur | pression | cholester | sucrer | 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 | oui | 31.00 | 3.00 | A | presence |
| 19 | 64.00 | masculin | A | 110.00 | 211.00 | A | C | 144.00 | oui | 18.00 | 2.00 | A | absence |

5. Conclusion

Voilà très brièvement la description de l'installation et de l'utilisation de l'add-in « sipina.xla ». Pour les férus d'informatique, le code source est consultable dans l'éditeur Visual Basic d'Excel. Les initiés constateront très vite que le programme est somme toute relativement simple. J'utilise le presse-papiers de Windows pour transférer les données. Le même principe est utilisé pour Tanagra. Ce schéma de fonctionnement est opérationnel avec Open & Libre Office, y compris sous Linux⁶.

⁶ <http://tutoriels-data-mining.blogspot.fr/2009/04/connexion-open-office-calc-sous-linux.html> pour le logiciel Tanagra.