# LES LOGICIELS ET L'ENSEIGNEMENT DE LA STATISTIQUE DANS LES DÉPARTEMENTS « STATISTIQUE ET INFORMATIQUE DÉCISIONNELLE » (STID) DES IUT

Gérard GRÉGOIRE, François-Xavier JOLLOIS, Jean-François PETIOT, Abdellah QANNARI, Serge SABOURIN, Philippe SWERTWAEGHER, Jean-Christophe TURLOT, Vincent VANDEWALLE, Sylvie VIGUIER-PLA[1]

**TITLE**

Software and statistics teaching in STID departments of IUT

**RÉSUMÉ**

Cet article fait le point sur les logiciels de statistique enseignés et utilisés dans les départements « STatistique et Informatique Décisionnelle » (STID) des Instituts Universitaires de Technologie (IUT). Quelles sont les raisons qui ont poussé les départements à choisir tel logiciel plutôt qu'un autre ? Pourquoi tel logiciel est-il apparu incontournable ? Quelle politique des départements vis-à-vis des logiciels gratuits ? Quelles sont les qualités et défauts des logiciels choisis ? Quelles pratiques pédagogiques pour l'enseignement de ces logiciels ? Quel ressenti des étudiants et des enseignants ? Quelle relation avec les éditeurs ? Quelle évolution à court terme ?

Pour essayer de répondre à ces questions, une enquête a été conduite auprès des enseignants des départements, avec un responsable par logiciel (les personnes citées sous le titre de l'article). Dans chaque département, une ou plusieurs personnes se sont chargées de répondre pour leur département aux divers questionnaires (cf. la liste des remerciements). Le questionnaire de base a été proposé par Sylvie Viguier-Pla. La coordination de l'enquête, la synthèse des contributions et la rédaction finale ont été réalisées par Jean-François Petiot et Gérard Grégoire.

***Mots-clés :*** *logiciels, enseignement de la statistique.*

[1] Les auteurs sont tous membres de départements « STatistique et Informatique Décisionnelle » (STID) de différents IUT : Gérard GRÉGOIRE (Grenoble, gerard.gregoire@iut2.upmf-grenoble.fr), François-Xavier JOLLOIS (Paris, francois-xavier.jollois@parisdescartes.fr), Jean-François PETIOT (Vannes, jean-francois.petiot@univ-ubs.fr), Abdellah QANNARI (Niort, abdellah.qannari@univ-poitiers.fr), Serge SABOURIN (Niort, srgsab@gmail.com), Philippe SWERTWAEGHER (Lisieux, Philippe.Swert@lisieux.iutcaen.unicaen.fr), Jean-Christophe TURLOT (Pau, jean-christophe.turlot@univ-pau.fr), Vincent VANDEWALLE (Roubaix, vincent.vandewalle@univ-lille2.fr), Sylvie VIGUIER-PLA (Carcassonne, viguier@cict.fr).

**Remerciements :** Pierre BARBILLON (Paris), Chantal BEDECARRAX (Grenoble), Delphine BLANKE (Avignon), Lydia BOUDJELOUD (Metz), Noëlle BRU (Pau), Thomas BURGER (Vannes), Jean-Yves CANDALEN (Lisieux), Alain CHOMEL (Carcassonne), Benoît COSSON (Vannes), François ELLAZAOUI (Paris), Gabriel FRAÏSSE (Carcassonne), Edith GABRIEL (Avignon), Philippe GARAT (Grenoble), François GAUTERO (Menton), Hervé GOLDFARB, (Lyon-Bron), Caroline HERMANN (Roubaix), Jean-Marc HUGONNET (Carcassonne), Yves LEMOINE (Metz), Tanguy LE NOUVEL (Vannes), Élisabeth LE SAUX (Vannes), Sabine LOUDCHER (Lyon-Bron), Alain LUCAS (Lisieux), Franck MARCHETTI (Metz), François MICHEL (Vannes), Michèle MOINE (Grenoble), Martine MOISI (Vannes), Souffien MOSBAH (Carcassonne), Florence NICOLAU (Menton), Jean-Michel POGGI (Paris), Olivier RENAULT (Grenoble), Adeline SAMSON (Paris), Saïd SERBOUTI (Roubaix), Rémi SERVIEN (Carcassonne), Nathalie VILLA-VIALANEIX (Carcassonne), Marlène VILLANOVA-OLIVER (Grenoble)

**ABSTRACT**

This paper provides an overview of the statistical software taught and used for teaching purposes in the STID departments of University Technological Institutes in France. What factors influenced the departments in their choice of computer software? What software characteristics were the deciding factors? What is the departments' policy on freeware? What are the advantages and disadvantages of the software currently used? How is the software used for teaching purposes? Are students and instructors happy with it? How are the relations with the software providers? Are any changes expected in the near future?

In order to answer these questions, the authors conducted a survey of instructors in the STID departments; each author was responsible for specific software. Each department appointed one or more representatives in charge of gathering the relevant information. Sylvie Viguier-Pla designed the basic questionnaire. Jean-François Petiot and Gérard Grégoire coordinated the survey, synthesized the contributions and produced the final report.

***Keywords:*** *software, statistical education.*

# 1 Introduction

## 1.1 Présentation des départements STID des IUT

L'enseignement de la statistique dans les IUT a débuté dès la création de ceux-ci en 1968 avec deux départements créés à Grenoble et à Paris, puis à Vannes en 1971. Huit autres départements ont été créés entre 1989 et 2000, à Pau, Niort, Metz, Roubaix, Carcassonne, Menton, Lyon-Bron et Lisieux. Le douzième département est né en Avignon en 2005. A l'origine, la spécialité était principalement tournée vers les applications à l'économie et à la gestion ; elle était intitulée « **S**tatistique, Etudes Economiques et **T**echniques **Q**uantitatives de Gestion » (STQ). Le Programme Pédagogique National (PPN) établi sous le contrôle de la Commission Pédagogique Nationale (CPN) a été profondément remanié en 1986 en repositionnant la spécialité, qui a pris le nom de « **S**tatistique et **T**raitement **I**nformatique des **D**onnées » (STID), vers tous les domaines d'application de la statistique, et en visant une double compétence statistique-informatique. Une nouvelle refonte du PPN a été mise en œuvre en 2005. Elle propose, comme dans les autres spécialités, des parcours différenciés aux étudiants visant une insertion professionnelle rapide (DUT à Bac+2 ou Licence Professionnelle à Bac+3 – parcours « court ») ou différée (master à Bac +5 – parcours « long »). La différence entre ces parcours peut être schématisée par « plus d'informatique » pour les parcours courts et « plus d'enseignements fondamentaux en mathématique et en statistique » pour les parcours longs. Ce nouveau programme a pris en compte l'évolution des métiers et des outils statistiques (logiciels) et informatiques. La maîtrise de logiciels professionnels, l'administration et l'exploitation des bases de données sont des points forts de la formation et il a ainsi été décidé en 2009 de souligner la dimension « informatique décisionnelle » de celle-ci par l'adoption de la dénomination « **ST**atistique et **I**nformatique **D**écisionnelle » qui conserve l'acronyme « STID ».

Les départements STID délivrent environ 500 DUT par an (478 en 2011) et ont diplômé environ 10 000 étudiants depuis leur création.

## 1.2 La place des logiciels dans le Programme Pédagogique National (PPN)

Le Programme Pédagogique National (PPN) se compose d'enseignements effectués par un enseignant ou un professionnel (environ 15 à 20% du volume des enseignements) dont le volume total sur les 4 semestres est de 1 620 heures, auxquelles s'ajoutent des projets tutorés (300 heures) et un stage de 10 semaines au minimum. Les 1 620 heures sont réparties en trois unités d'enseignement (UE) :

- Statistique (545h) ;
- Outils scientifiques (515h : mathématique 120h, informatique 375h, data-mining 20h) ;
- Environnement économique et communication (économie, gestion, expression-communication, anglais 560h).

A ces trois UE s'ajoute l'UE « Projets tutorés et stages » aux semestres 2 et 4. Dans les départements STID (ce n'est pas toujours le cas dans les autres spécialités) le stage de 10 semaines est placé à la fin du semestre 4.

La formation STID vise à une double compétence statistique et informatique.

Dans le domaine de l'informatique, la formation dispensée vise d'une part à donner de solides bases en informatique générale (environnement informatique, bureautique, algorithmique et programmation), d'autre part à doter l'étudiant d'une compétence affirmée dans le domaine de la gestion des données : un diplômé STID doit être capable en particulier de concevoir et d'administrer une base de données. Les enseignements relatifs à cette partie de la formation constituent 320 des 375 heures dédiées à l'informatique.

### 1.2.1 Les logiciels et les modules « logiciels spécialisés »

Le complément de ces 320 heures, soit un volume de 55h, est dévolu à un enseignement de logiciels spécialisés de statistique (dans le PPN, module « Logiciels spécialisés ») réparti sur les trois premiers semestres. Il s'agit d'un enseignement visant à faire des diplômés des spécialistes des logiciels de statistique utilisés dans le monde professionnel. La totalité des 12 départements consacrent une grande partie de ce volume d'heures à un enseignement du logiciel SAS complété le plus souvent par celui de R, de SPSS, de MAPINFO.

Notons aussi que les volumes horaires donnés ci-dessus constituent une base qui sera modifiée selon le parcours choisi par l'étudiant. Si l'étudiant se destine à une insertion professionnelle rapide (à la sortie du DUT ou après une Licence Professionnelle), un module de 35 heures sur les logiciels spécialisés sera ajouté aux 55 heures du tronc commun, dans le but d'accentuer la professionnalisation de la formation.

### 1.2.2 Les logiciels et les modules d'enseignement de la statistique

Il est à noter que le module « logiciels spécialisés » ne constitue qu'une partie des enseignements où l'étudiant pratique les logiciels permettant de faire du traitement statistique. En effet, tous les enseignements de statistique utilisent des logiciels, dans une proportion variant de 10% à 40% du volume de l'enseignement concerné, aussi bien pour la compréhension et l'assimilation des méthodes que pour la mise en œuvre et l'acquisition du savoir-faire sur des données réelles ou artificielles. Les modules concernés sont les suivants :

- statistique descriptive (semestre 1) ;
- études statistiques (participation à une étude – semestre 1) ;
- études statistiques, chroniques et simulation (semestre 2) ;
- estimation et tests (semestre 3) ;
- régression et analyse de la variance (semestres 3 et 4) ;
- sondages (semestre 4) ;
- analyse des données (semestres 3 et 4) ;
- tests non paramétriques (semestre 4) ;
- data mining (semestre 4).

Le « data mining » est souvent considéré comme un ensemble de méthodes à l'interface de la statistique et de l'informatique. Le PPN STID a choisi de mettre l'accent sur les méthodes d'aide à la décision : l'analyse discriminante, la régression logistique, les arbres de décision.

Pour les enseignements cités ci-dessus, chaque département a recours à plusieurs des logiciels répertoriés dans le tableau donné plus bas.

Enfin, les projets tutorés et le stage constituent des opportunités importantes d'utilisation des logiciels. En particulier, les logiciels proposés pourront être pour l'étudiant un des arguments de choix entre plusieurs propositions de stage. En définitive, ces logiciels sont très présents dans l'ensemble de la formation en statistique.

TABLEAU 1 – *Les logiciels utilisés pour l'enseignement de la statistique dans les départements STID. Les logiciels à gauche de la double barre verticale sont des logiciels généralistes, ceux de droite des logiciels dédiés à des domaines particuliers de la statistique.*

| | **SAS** | **Excel** | **R** | **SPSS** | **Xlstat** | **Statgra-phics** | **SPAD** | **Le Sphinx** | **MAPINFO** | **Tanagra** |
|---|---|---|---|---|---|---|---|---|---|---|
| Avignon | X | X | X | | | | | | | |
| Carcassonne | X | X | X | | X | | X | X | X | |
| Grenoble | X | X | X | X | | | X | X | X | X |
| Lisieux | X | | X | X | | | X | | X | |
| Lyon | X | X | X | X | | | X | X | X | X |
| Menton | X | X | X | | X | | | X | | |
| Metz | X | X | X | X | | | | | | |
| Niort | X | X | X | X | | X | X | X | | |
| Paris | X | X | X | X | | | X | X | | X |
| Pau | X | X | X | | | | | X | | |
| Roubaix | X | X | X | | | | | X | | X |
| Vannes | X | X | X | | | X | X | | X | |

# 2 Les logiciels incontournables

## 2.1 EXCEL

Le logiciel Excel de la suite Office de Microsoft est, comme chacun le sait, un tableur. Il possède un certain nombre de fonctions statistiques, et on peut étendre ses capacités sur ce plan en lui adjoignant des macros mises à disposition par Microsoft, telles que « Utilitaire d'analyse » et « Utilitaire d'analyse-VBA », ou disponibles sur internet.

Parmi les 12 départements STID, pour 10 d'entre eux, Excel joue un rôle dans le dispositif d'enseignement de la statistique (et 11 si l'on compte un département qui recourt à « Open Office Calc », logiciel libre de conception voisine de celle d'Excel). Un seul département n'a recourt ni à Excel, ni à un logiciel libre du même type.

**Raisons de l'utilisation de ce logiciel**

- Le logiciel est présent quasiment partout dans le monde professionnel, entreprises ou administrations.
- Le logiciel est parfois le seul disponible dans l'entreprise pour effectuer des traitements statistiques.
- Les données circulent souvent au format Excel ou sont disponibles dans une base de données comme Access. Le traitement statistique peut s'effectuer sans faire passer les données par une phase de conversion. De même la saisie directe des données est facile.
- Les étudiants sont en général à l'aise avec Excel à leur arrivée en STID, ou le deviennent très rapidement. Le logiciel est facile à utiliser, produit des graphiques attractifs de façon immédiate. Lorsque les étudiants commencent à avoir un peu d'expérience, le lien direct avec VBA leur permet de produire des rapports automatisés.
- D'un point de vue pédagogique, Excel présente l'intérêt de permettre une décomposition fine des procédures statistiques, qui peut être exploitée pour l'acquisition des méthodes et résultats de base de la statistique. L'étudiant voit en même temps les données, les calculs intermédiaires, les graphiques, les résultats. Cela n'est pas possible avec un logiciel « clique-bouton » et ne nécessite pas de formation à la programmation, formation que les étudiants ne possèdent pas en arrivant en 1re année.

**Modalités d'enseignement**

- L'enseignement d'Excel proprement dit est en général dispensé par les enseignants d'informatique.
- Tous les départements qui utilisent Excel dans les enseignements de statistique, l'utilisent en particulier pour l'enseignement de la statistique descriptive en S1, pour un volume d'heures variant entre 25% et 50% du volume de TD.
- Au gré des départements, et avec une importance variable, on trouve le recours à Excel dans la plupart des secteurs de la statistique enseignée en STID : études statistiques (3 dépts.), statistique inférentielle, analyse de la variance, modèle linéaire

(4 dépts.), probabilités, échantillonnage-simulation (4 dépts.), analyse des données (1 dépt.).

- Le logiciel est utilisé pour l'apprentissage de la statistique par expérimentation : l'étudiant redécouvre par lui-même, par le biais de simulations, tel ou tel résultat de probabilités ou de statistique. On a recours à Excel pour illustrer, sur des exemples simples de données, des procédures présentées en cours. Ce logiciel permet très simplement de tirer un échantillon aléatoire dans une base de sondage, selon un plan de sondage, en population finie. Il permet aussi d'effectuer, dans un cadre limité, un traitement sur des données réelles, notamment dans le cadre des projets tutorés de 1re année. Enfin, il sert aussi de boîte à outils rapide (lecture de tables, calculs d'indicateurs statistiques, construction de graphes…).

**Retour d'expérience**

- Les atouts d'Excel sont clairement le fait que ce logiciel est présent quasiment partout dans le secteur professionnel, qu'il est facile d'utilisation et qu'il présente un grand intérêt comme outil d'apprentissage basé sur la simulation.
- Points négatifs : le logiciel est avant tout un tableur et pas un logiciel dédié à la statistique. On observe des lacunes à des niveaux divers : pas de possibilité de construire de véritables histogrammes avec des classes d'amplitudes différentes, pas de boîtes à moustaches (box-plot). Les traitements statistiques offerts restent assez limités, même si des macros peuvent être trouvées pour enrichir les fonctionnalités. On peut être conduit à effectuer des tâches particulièrement répétitives. L'aide a été traduite en français de façon très approximative et la dénomination des fonctions semble souvent peu naturelle.
- Les étudiants apprécient en général que la statistique soit enseignée en ayant recours à Excel. Ils sont souvent très à l'aise dans l'utilisation de ce logiciel. Leur savoir-faire dans l'utilisation de ce logiciel couplé avec leurs connaissances en statistique est apprécié des entreprises dans lesquelles ils font leur stage.

**Bilan et perspectives**

Excel fait actuellement partie intégrante du dispositif pédagogique d'enseignement de la statistique dans les départements STID. Le logiciel est surtout utilisé en 1re année, même si certains départements y recourent aussi en 2e année. Il joue clairement un rôle dans le dispositif d'apprentissage pour l'acquisition et la compréhension des méthodes et résultats fondamentaux de la statistique. Il permet aussi de faire du traitement statistique sur des données à un moment du processus d'apprentissage où le recours aux logiciels « clique-bouton » n'est pas souhaitable, et où les étudiants ne sont pas encore suffisamment formés à la programmation pour utiliser R ou SAS.

Dans la plupart des départements, on n'envisage pas de réduire le recours à Excel dans l'enseignement de la statistique. Il a un rôle bien identifié dans le dispositif d'enseignement de la statistique, aux côtés de logiciels du monde professionnel spécifiquement dédiés à la statistique tels que SAS, SPSS, ou de logiciels du « monde libre » tels que R, pour ne citer que les plus utilisés.

## 2.2 SAS

Le logiciel SAS est incontournable dans le domaine de la statistique et de l'informatique décisionnelle. Il est enseigné depuis de nombreuses années dans tous les départements STID. Il permet d'appréhender l'ensemble des étapes d'un processus décisionnel, de l'importation des données à partir d'un système dédié jusqu'à la mise en forme et la diffusion des résultats.

**Raisons de l'utilisation de ce logiciel**

- SAS est très répandu dans le monde professionnel, particulièrement dans le monde de la santé (laboratoires pharmaceutiques, hôpitaux, CRO, Sécurité Sociale, mutuelles…), mais aussi dans le milieu bancaire, les assurances, les sociétés d'enquête, d'études de marché, les grandes industries, les administrations… Certaines offres d'emploi comportent la mention « la connaissance de SAS est impérative ».
- Le traitement statistique offert par SAS est particulièrement riche et complet, voire même exhaustif. Il couvre la totalité des domaines particuliers de développement des méthodes statistiques de l'économétrie à la biostatistique, en passant par le contrôle de qualité, le marketing, les SIG…
- SAS est particulièrement « robuste » et peut traiter des jeux de données très volumineux (plusieurs millions d'individus).
- Ayant son langage propre, il est parfois difficile pour des étudiants peu formés à ce logiciel de créer des macros SAS en stage et/ou en situation d'emploi. Par contre, il permet de mieux maîtriser les calculs effectués, ce qui est un plus pédagogique.
- SAS met à disposition des enseignants, et des étudiants pendant leurs études, une version personnelle. Les étudiants peuvent travailler chez eux avec cette version comme s'ils étaient dans une salle de l'IUT. Enfin SAS propose aussi un dispositif pour qu'un étudiant puisse utiliser le logiciel en stage dans une société qui n'en dispose pas.

**Modalités d'enseignement**

- Quelques départements commencent à enseigner SAS dès la première année, sur la partie importation et gestion des données. En seconde année, le logiciel est utilisé pour divers enseignements en statistique (tels que l'analyse de la variance, la classification et l'analyse de données, les tests d'hypothèse…). La partie graphique est aussi enseignée dans plusieurs départements.
- L'enseignement a très souvent lieu en salle d'ordinateurs, sous une forme qui combine apprentissage et pratique. Selon les départements et le contenu des cours, il est effectué par des mathématiciens, des statisticiens ou des informaticiens. Une partie de l'enseignement peut être faite dans le cadre des modules « logiciels spécialisés ». Cela peut concerner par exemple la gestion de données, la programmation, l'édition de rapports (reporting), sujets qui pourront être enseignés par des informaticiens, ou des mathématiciens qui ont investi dans le domaine. Les procédures statistiques (pour la statistique inférentielle de base, la régression, l'analyse de variance, l'analyse de données…) peuvent être ensuite présentées par les enseignants de statistique dans le cadre des enseignements concernés.

– L'évaluation est basée la plupart du temps sur un examen final, en temps limité, parfois sur un ordinateur, parfois sur papier. Dans quelques départements, les étudiants ont des travaux pratiques faisant l'objet d'une évaluation.

**Retour d'expérience**

– Les étudiants montrent parfois au départ un intérêt limité pour ce logiciel, mais qui devient grandissant lorsqu'ils comprennent son importance dans la plupart des stages qui leur sont proposés. Toutefois, c'est un enseignement difficile du fait du langage spécifique et différent des langages de programmation classiques. Les étudiants ont parfois du mal à intégrer certains aspects.

– Par contre, les entreprises ayant en stage ou recrutant des étudiants du DUT STID sont en grande majorité très satisfaites des compétences de ceux-ci dans l'utilisation de ce logiciel. C'est un point très positif pour notre formation, qui est ainsi très appréciée du monde professionnel, qui exige souvent ce profil pour des emplois de « programmeurs statistiques ».

**Bilan et perspectives**

Malgré quelques difficultés dans l'enseignement et la réticence des étudiants au début, les enseignants sont globalement satisfaits de ce logiciel. Le principal problème est le manque de temps pour montrer l'ensemble des possibilités de SAS. Mais, ayant acquis les bases et compris la logique, les étudiants arrivent aisément à se former par eux-mêmes lorsqu'ils sont confrontés à des aspects qu'ils n'ont pas étudiés en cours.

Etant donné l'implantation de SAS dans le monde professionnel, les départements STID ne peuvent pas se passer de former leurs étudiants à son utilisation. Cependant les départements sont unanimes à considérer que le prix des licences SAS fait peser un poids trop lourd sur leur budget. D'autre part, des logiciels viennent maintenant concurrencer SAS : un logiciel gratuit comme R rivalise de plus en plus avec SAS sur le plan de l'exhaustivité des méthodes statistiques proposées, en particulier dans les développements méthodologiques récents.

## 2.3 R

R est un logiciel statistique qui permet la lecture, la manipulation et le stockage de données. La grande majorité des méthodes statistiques actuelles y sont présentes par défaut ou au sein de « packages » dont la liste est en constante évolution. Les principaux facteurs qui expliquent son importance actuelle sont sa gratuité (licence GNU-GPL), sa fiabilité et sa disponibilité sous la plupart des systèmes d'exploitation (Windows, Mac OS, Linux, Unix). Initialement conçu pour illustrer l'enseignement de la statistique, R a connu une croissance exponentielle pendant les quinze dernières années dans le monde académique. Son développement actuel lui permet de rivaliser avec la plupart des logiciels payants utilisés dans les entreprises.

**Raisons de l'utilisation de ce logiciel**

– R est enseigné dans les 12 départements STID pour mettre en œuvre les méthodes statistiques étudiées. Il ne présente aucun problème d'accessibilité puisqu'il peut être téléchargé gratuitement sur n'importe quelle machine.

- D'un point de vue pédagogique, la programmation permet aux étudiants de mieux appréhender les différentes étapes des méthodes statistiques employées contrairement aux logiciels de type « clique-bouton » qui apparaissent parfois comme des boîtes noires.
- De plus, compte tenu de son évolution actuelle, on peut penser que sa maîtrise va faire rapidement partie des compétences utiles à l'insertion professionnelle des étudiants.

**Modalités d'enseignement**

L'utilisation de R est assez variable selon les départements. Elle va de la simple initiation à une utilisation quasi-systématique.

- L'enseignement est principalement dispensé par des statisticiens pour illustrer les notions étudiées dans leur discipline. Par exemple, dans le module « technique de simulation », R permet de générer de nombreux types de variables aléatoires, de calculer facilement les indicateurs utiles et de faire les graphiques associés.
- Il est aussi parfois utilisé en projet tutoré.
- Dans certains départements il est utilisé par des informaticiens en cours de programmation objet.

Son utilisation se fait en général sous forme de TP, voire de TD-TP. L'évaluation repose sur un examen final sur ordinateur ou sur papier et parfois aussi sur des TP notés.

**Retour d'expérience**

R est généralement moins apprécié par les étudiants que les logiciels à base de menus déroulants. D'une part, sa prise en main apparaît plus complexe du fait de l'apprentissage d'un langage dans un volume d'enseignement souvent insuffisant pour rendre les étudiants autonomes dans sa manipulation. D'autre part, le logiciel est encore peu implanté dans le monde professionnel et les entreprises sont encore peu demandeuses de cette compétence. Les étudiants ne considèrent souvent pas la connaissance de R comme un atout sur le marché du travail. Cependant, quand les étudiants proposent l'utilisation de ce logiciel durant leur stage, les retours sont généralement très positifs. La grande diversité des méthodes disponibles suscite toutefois l'intérêt des étudiants. Les plus intéressés explorent ainsi de nouvelles fonctionnalités statistiques dans le prolongement des notions étudiées en cours en fonction des besoins rencontrés en stage.

**Bilan et perspectives**

Une fois l'apprentissage des bases dispensé, R peut être utilisé tout au long de la formation pour illustrer les différentes méthodes statistiques enseignées dans les départements STID. Actuellement peu utilisé en milieu professionnel, on peut s'attendre à un développement rapide de son utilisation, notamment du fait de sa totale gratuité et du développement de modules qui le rendent plus accessible. Citons, par exemple, Rcommander, une interface à base de menus déroulants qui facilite la prise en main du logiciel, Rexcel qui simplifie la communication entre R et Excel ou encore Odfweave qui permet l'automatisation de la production de rapports.

On peut s'interroger sur l'équilibre qui sera réalisé dans les années à venir entre l'enseignement de ce logiciel prometteur et celui des logiciels de statistique du monde

professionnel, la maîtrise de ces derniers restant actuellement incontournable pour l'insertion professionnelle des étudiants.

# 3 Les logiciels courants

## 3.1 SPSS

Ce logiciel est un logiciel généraliste, bien qu'à l'origine orienté vers la statistique en sciences sociales (Statistical Package for Social Sciences). Aux Etats-Unis et en Grande-Bretagne, son utilisation est largement répandue dans le monde professionnel et ce dans les secteurs les plus variés. Il rencontre un succès moins large en France mais occupe cependant une place importante. Il couvre tous les champs de la statistique et il est basé sur l'utilisation de menus déroulants ; sa prise en main est rapide.

**Raisons de l'utilisation de ce logiciel**

Ce logiciel est enseigné dans sept des douze départements STID pour trois raisons essentielles :

- SPSS est un logiciel professionnel (récemment racheté par IBM) largement utilisé à l'étranger (Etats-Unis, Grande-Bretagne, etc.) et qui se diffuse en France (il est présent dans des entreprises partenaires des IUT). Il a de grandes capacités de calcul et il est donc intéressant qu'il figure dans la liste des logiciels étudiés pour la recherche de stage et, à terme, pour la recherche d'emploi des étudiants. Certains départements ont un lien étroit avec le service « Formation » de SPSS .
- Du point de vue pédagogique, ce logiciel offre l'avantage d'une utilisation facile, sans apprentissage informatique préalable (barre des tâches et menus déroulants avec des procédures prédéfinies) ce qui en autorise l'utilisation dès le début de la formation après une prise en main rapide. Il permet de mettre en application les méthodes statistiques et de servir de support à leur interprétation sur des données concrètes. Pour l'automatisation des tâches, ou la réalisation de celles qui ne figurent pas dans le pack de base, il peut être utilisé en programmant grâce à un éditeur de syntaxe (langage ressemblant à celui de SAS avec procédures et sous-commandes associées).
- Les sorties graphiques sont de très bonne qualité et l'éditeur de graphiques permet des mises en forme personnalisées très appréciées pour les restitutions audiovisuelles ou imprimées.

**Modalités d'enseignement**

Les modalités d'enseignement sont de deux types selon les départements :

- Trois départements ont fait le choix de mettre en application les enseignements des modules de statistique descriptive, séries chronologiques, estimation et tests, régression linéaire et analyse de la variance, et analyse des données avec SPSS sous forme de TD ou TP sur ordinateur tout au long du cursus (ce qui représente de 20h à 40h en première année et autant en seconde année).

– Quatre départements ciblent la formation en seconde année pour les étudiants qui utiliseront SPSS lors de leur futur stage. Dans ce cas le volume horaire est de l'ordre de 10h à 20h.

**Retour d'expérience**

– Aspects positifs : facilité d'apprentissage, possibilité pour l'enseignant d'axer le commentaire sur les aspects statistiques sans parler d'informatique, possibilité d'enregistrer son travail pour le reprendre ultérieurement. Possibilité d'avoir une version en prêt gratuit durant le stage. La capacité qu'ont les étudiants à utiliser SPSS est appréciée des entreprises possédant ce logiciel.

– Aspects négatifs : difficulté au début de se rappeler le chemin dans les menus déroulants (propre à tous les logiciels « clique-bouton »), documentation en français mal traduite et source d'erreurs (il est recommandé d'utiliser la documentation en anglais). La licence est chère et il n'existe pas de licence gratuite pour les étudiants.

– Les étudiants doivent réfléchir sur les sorties (trier la bonne information) sans avoir à passer du temps sur une étape « programmation » (comme sous SAS), ce qui a l'avantage de permettre une pédagogie plus concentrée sur les aspects statistiques.

– Dans le cadre de la formation en apprentissage, les étudiants peuvent se trouver en situation d'avoir à utiliser en entreprise ce logiciel alors que son enseignement n'est pas encore très avancé. L'expérience montre que les étudiants parviennent généralement à s'autoformer sans trop de difficultés.

**Bilan et perspectives**

– Ce logiciel reste un complément intéressant à l'apprentissage de SAS (utilisation « clique-bouton » d'un logiciel professionnel).

– La licence est chère et l'utilisation du logiciel pourrait être réduite, voire abandonnée, si l'éditeur ne fait pas de geste en direction des départements qui utilisent ce logiciel, notamment la délivrance d'une licence gratuite pour les étudiants et les enseignants.

– Le développement de logiciels gratuits tels que R, Tanagra, etc., devrait sans doute inviter SPSS à revoir sa politique commerciale vis-à-vis de l'enseignement de son logiciel.

## 3.2 SPAD

Le logiciel SPAD est essentiellement orienté vers l'analyse des données et l'aide à la décision. Il permet de couvrir l'ensemble du processus d'aide à la décision, du management des données à la prise de décision, en passant par un choix varié d'outils d'analyse exploratoire des données. Ce logiciel procède par choix de méthodes qui sont enchaînées sur la base d'un menu. Près de deux départements sur trois sont familiers de l'enseignement de ce logiciel.

**Raisons de l'utilisation de ce logiciel**

– SPAD permet d'illustrer clairement la phase statistique du processus de data mining : techniques descriptives et d'analyse des données pour l'appropriation et la validation

des données, puis méthodes d'aide à la décision (méthodes d'analyse discriminante, méthodes de segmentation, etc.), en amont des outils de management des données.

- La qualité des graphiques et les possibilités de leur personnalisation (couleurs, tracés, zooms, etc.) au moyen de manipulations simples sont un atout fort de ce logiciel.
- La disponibilité des résultats des analyses sous forme de fichiers EXCEL permet des enchainements aisés avec d'autres méthodes que l'on trouvera dans d'autres logiciels comme SAS ou SPSS.
- La facilité d'utilisation : même s'il ne s'agit pas d'un logiciel « clique-bouton », l'enseignant n'a pas besoin de consacrer des séances spécifiques à l'apprentissage du logiciel et l'étudiant peut être ainsi entièrement concentré sur « la méthode ».
- La pertinence des résultats fournis par le logiciel.

**Modalités d'enseignement**

SPAD ne nécessite pas d'apprentissage spécifique ; c'est pourquoi il est employé dans le cadre de TD (les résultats fournis sont en très bonne adéquation avec le développement du cours). Les départements possédant ce logiciel l'utilisent en TD et en TP conjointement avec R, par exemple, dans les modules analyse des données 1 (semestre 3), analyse des données 2 (semestre 4) et data mining (aide à la décision, semestre 4). Il est ensuite largement utilisé par les étudiants dans le cadre des projets de seconde année, sans véritable difficulté d'utilisation.

**Retour d'expérience**

- Points positifs : les étudiants apprécient sa facilité d'utilisation, le fait qu'on ait accès à une suite de procédures sous forme d'icônes permettant de retracer la séquence des traitements effectués. Le logiciel est particulièrement adapté aux enseignements d'analyse de données et de data mining. Il est mis gratuitement à disposition des étudiants dans le cadre de leur stage. Le responsable de l'enseignement peut disposer d'une version personnelle, de même que l'étudiant pendant son cursus en STID.
- Point faible : ce logiciel est plus utilisé dans les services de recherche que dans les systèmes d'information des entreprises, où il est peu présent.

**Bilan et perspectives**

Le logiciel SPAD trouve sa genèse dans les travaux d'une équipe de recherche qui a joué un rôle très important dans l'avènement de l'analyse des données « à la française ». Si, bien entendu, ce logiciel s'est fortement développé depuis, il conserve certaines spécificités qui en font un outil pédagogique original et de qualité. On dispose maintenant sous R de bibliothèques de programmes calquant les solutions SPAD en analyse des données, ce qui peut à terme porter préjudice à ce dernier dans le cadre de l'enseignement en STID.

## 3.3 MAPINFO

Mapinfo est un logiciel de traitement de l'information géographique. Il permet la représentation de données statistiques sur des cartes mais comporte aussi un volet gestion de bases de données géolocalisées et programmation. Le fait qu'il ait été pionnier dans le domaine explique en partie sa position privilégiée dans un créneau de logiciels qui connaît actuellement une forte croissance.

**Raisons de l'utilisation**

La représentation de données statistiques sur des cartes est la première motivation de cet enseignement. Le choix du logiciel MAPINFO pour cela vient de son caractère pionnier, de sa grande convivialité dans son utilisation et du fait qu'il est assez présent dans le monde professionnel, notamment dans les collectivités locales. Avec l'expérience, il s'avère que la composante géographique prend une importance croissante dans le monde professionnel. En effet, on la retrouve pour la visualisation de données démographiques, sociales, d'économie territoriale, d'aménagement du territoire, de réseaux (voies de circulation, eau, téléphonie, gaz, électricité, eaux usées…), de logistique (transports, secours, dépannages…), d'environnement, de météorologie, et de bien d'autres exemples encore. Au-delà de la simple représentation, la facilité de lecture d'une situation en fait un outil décisionnel pour l'entreprise.

**Modalités d'enseignement**

On distingue deux aspects dans l'enseignement de MAPINFO :

- Le plus généralisé est celui de l'apprentissage de la manipulation du logiciel pour la mise sur carte de données statistiques (Carcassonne, Grenoble, Menton, Vannes). Cet aspect est le plus souvent vu en 1re année dans le cadre des « logiciels spécialisés ». Le contenu du cours comprend la présentation générale de ce qu'est une carte, avec ses différentes représentations possibles, et la manipulation du logiciel. Il s'accompagne d'applications qui amènent à aborder d'autres enseignements comme l'économie descriptive, la géographie, la conduite de projet… Cet enseignement représente de 6 à 20 heures de TD/TP. Les étudiants peuvent aussi être confrontés à l'utilisation de MAPINFO dans le cadre des projets tutorés.
- Le second aspect est celui d'une prise en main plus approfondie du logiciel, avec gestion de bases de données géolocalisées, et programmation. Ce second aspect est enseigné dans les départements qui sont aussi porteurs de licences professionnelles spécialisées dans la géographie (Carcassonne, Grenoble). Cela implique un investissement plus important dans les logiciels de cartographie, notamment la programmation avec MAPBASIC, mais aussi l'enseignement d'autres logiciels autour de cette cartographie, comme celui de Mapserver, Postgre SQL, … qui permettent de programmer pour la manipulation de cartes, ainsi que le développement de l'interactivité des cartes sur internet. Jusqu'à 100 heures peuvent être consacrées à cet enseignement.

**Retour d'expérience**

- Le temps imparti (entre 6 et 20h selon les départements en STID, jusqu'à 100h en LP) pour cet enseignement est souvent jugé insuffisant, et la disponibilité des logiciels aussi, car le coût du logiciel empêche son installation généralisée sur les sites d'enseignement, et donc une utilisation intensive.
- Le retour de la part des étudiants est très positif, surtout de ceux de licence professionnelle, que le contact avec l'entreprise a sensibilisés au besoin croissant de la connaissance de ce type de logiciel.
- Les enseignants sont très réactifs aux nouveautés autour de ce logiciel et les intègrent dans leurs projets d'évolution de leur enseignement.

**Bilan et perspectives**

- L'enseignement de MAPINFO n'est pas généralisé dans les départements STID, puisqu'il concerne une petite moitié de l'ensemble des sites d'enseignement. Son coût très élevé rapporté au peu d'heures d'enseignement consacrées ne le destine pas à une expansion.
- En revanche, les offres d'emploi demandant de plus en plus souvent la connaissance de Systèmes d'Information Géographiques (SIG), on peut envisager une généralisation d'un enseignement de ce type. Certains départements qui enseignent MAPINFO enseignent aussi d'autres logiciels de cartographie, et regardent vers les logiciels libres.
- L'atout de MAPINFO est d'être assez présent dans l'entreprise, ce qui donne un avantage à l'étudiant qui le connaît. Cependant, l'expansion à la fois de l'utilisation des SIG et des éditeurs de ces logiciels laisse largement la place à des produits concurrents.

# 4 Autres logiciels

Les autres logiciels cités ici sont en général peu utilisés pour des raisons variées, soit parce qu'ils sont très récents et encore peu connus, soit parce qu'ils sont jugés comme non suffisamment implantés en entreprise, soit parce qu'ils sont considérés comme trop spécialisés et non nécessaires en STID, soit enfin parce que l'éventail des logiciels enseignés dans le département paraît suffisant pour répondre aux objectifs pédagogiques. Un logiciel particulier peut aussi avoir été choisi dans un département parce que promu par un enseignant qui en avait une expérience approfondie. Les informations recueillies sur ces logiciels sont donc beaucoup plus parcellaires que celles obtenues pour les logiciels précédents et leur présentation plus rapide.

## 4.1 Logiciels généralistes : XLSTAT, STATGRAPHICS, MINITAB, JMP

**XLSTAT**

Xlstat est un logiciel Microsoft. Il s'agit d'un « add-on », qui complète Excel pour faire de la statistique. Il hérite donc des qualités d'Excel et lui rajoute beaucoup de fonctions statistiques qui lui faisaient défaut. Outre les outils habituels de statistique descriptive, sont présents entre autres les outils classiques de statistique inférentielle, les méthodes de régression (OLS, PLS), l'analyse de la variance, les méthodes factorielles, les outils d'analyse et de prévisions des séries chronologiques, etc. Xlstat couvre donc un champ suffisamment large pour le programme de statistique de STID. Deux départements l'utilisent en statistique descriptive et inférentielle, en simulation, régression, analyse de variance et en analyse des données. Les étudiants apprécient sa facilité d'utilisation et la clarté de ses sorties.

**STATGRAPHICS**

Statgraphics permet de réaliser aisément des analyses graphiques et statistiques. Deux départements STID utilisent ce logiciel, l'un depuis environ vingt ans et l'autre depuis la rentrée 2011/2012. Il est intégré dans l'enseignement pour illustrer et appliquer les méthodes statistiques étudiées : statistique descriptive (univariée et bivariée), modèle linéaire

(régression et analyse de la variance) mais aussi pour son module graphique qui propose un large éventail de représentations. En dehors des séances de travaux pratiques, les étudiants disposent du logiciel en libre service pour approfondir leurs connaissances ou réaliser les projets tutorés. C'est un logiciel très pédagogique assez simple d'utilisation et qui semble intéresser les étudiants. Néanmoins, son absence dans les entreprises en limite son usage au cadre universitaire.

**MINITAB**

Développé en 1972 à la Pennsylvania State University, Minitab (Minitab Inc.) est un logiciel de statistique offrant une panoplie de méthodes d'analyse statistiques allant de la statistique élémentaire à des traitements plus avancés (multivariés, modèle linéaire, survie…) en passant par la Maitrise Statistique des Procédés (MSP). Le choix de ce logiciel est dicté par sa facilité d'utilisation (il ne nécessite quasiment pas d'apprentissage), son caractère pédagogique et la présence d'un module spécifique MSP. Un seul département STID intègre Minitab dans son enseignement pour illustrer et mettre en œuvre les méthodes et outils statistiques étudiés. Outre les statistiques descriptives, Minitab est largement utilisé en régression et analyse de la variance (près d'un quart des enseignements), en MSP et plans d'expériences. C'est un outil pédagogique et performant pour l'apprentissage de la statistique en particulier sur de petits fichiers.

**JMP**

JMP est un logiciel commercialisé par SAS Institute et, de ce fait, bénéficie des mêmes conditions d'installation (version gratuite pour les étudiants et, selon la licence, possibilité de l'installer sur tous les ordinateurs sans restriction du nombre d'utilisateurs).

Un département a en projet de l'enseigner dès l'année prochaine. L'intérêt de ce logiciel, outre sa licence et ses liens forts avec SAS (possibilité d'importer et d'exporter vers SAS, etc.), est sa grande convivialité, sa facilité d'utilisation (pas de programmation obligatoire ; conçu à l'origine pour le Mac, il est très visuel) et l'étendue des méthodes présentes (propres et bien présentées, avec des graphiques utiles) bien adaptée au programme de l'enseignement de statistique en STID.

## 4.2 Logiciels dédiés : Sphinx, Arcgis, Lime survey, Tanagra, Knime

**LE SPHINX**

Le Sphinx est un logiciel de gestion d'enquêtes créé il y a 25 ans. Il est utilisé par la majorité des départements car il dispose d'une interface intuitive qui permet de créer facilement des questionnaires et d'en faire le traitement. Il trouve ainsi sa place dans les projets des étudiants, notamment en première année. La nécessaire formalisation des questions en utilisant Le Sphinx obligent ceux-ci à une réflexion qui aurait été parfois moins aboutie lors de la confection d'un questionnaire avec un simple éditeur de textes. L'option « Scanner » permet la saisie automatique de grandes quantités de questionnaires. L'option « Enquête Web » permet la mise en ligne facile des questionnaires et l'option « Lexica » permet aux étudiants de mettre en pratique l'analyse textuelle.

**ARCGIS**

L'utilisation de ARCGIS est réservée aux départements STID qui proposent une licence professionnelle en traitement des données géographiques (Carcassonne et Grenoble). Il

représente un complément à l'enseignement de MAPINFO. Les deux logiciels sont à peu près comparables quant à l'étendue de leurs fonctionnalités, mais certaines de ces fonctionnalités sont plus faciles avec l'un, et d'autres avec l'autre. ARCGIS permet de faire de la statistique spatiale (des méthodes d'interpolation sont implémentées). Son enseignement moins fréquent en DUT reflète l'utilisation moins répandue que celle de MAPINFO dans le monde professionnel.

**LIME SURVEY**

Lime Survey développé à l'origine sous le nom de PHPSurveyor en 2003 est un logiciel libre (sous licence GNU GPL). Ce logiciel permet de mettre en ligne des questionnaires à partir d'un navigateur Web. Les étudiants de plusieurs départements STID mettent en œuvre cette interface de sondage pour des enquêtes d'envergure nationale ou plus avec la possibilité de grosses volumétries de données. L'association STID France l'a utilisé lors d'une enquête nationale sur l'usage des moyens audiovisuels dans le supérieur pour le compte du ministère. Les données sont stockées dans une base de données MySQL.

Cette plateforme d'enquête est solide (MySQL) et d'installation simple, elle permet de gérer simultanément plusieurs enquêtes. La typologie des questions est très riche ainsi que les possibilités de contrôle. Le cheminement à l'intérieur d'un questionnaire est possible grâce à des conditions. Les choix et la création des modèles permettent une bonne souplesse dans la construction de l'interface. Il est possible et aisé de copier et de modifier un questionnaire. Lime Survey permet de publier une enquête d'une manière ciblée par l'envoi de mails ou ouverte par publication du lien. L'accès aux résultats est sécurisé. Les étudiants implantent parfois ce logiciel sur leur lieu de stage.

**TANAGRA**

Tanagra est un logiciel gratuit et open source de Data Mining écrit par Rico Rakotomalala du laboratoire ERIC (Université Lyon 2 Lumière). Il implémente une série de méthodes de fouilles de données issues du domaine de la statistique exploratoire, de l'analyse de données, de l'apprentissage automatique et des bases de données. Tanagra offre aux chercheurs et aux étudiants une plate-forme de Data Mining facile d'accès, respectant les standards des logiciels du domaine, notamment en matière d'interface et de mode de fonctionnement, et permettant de mener des études sur des données réelles et/ou artificielles. Bien qu'étant beaucoup plus limité (en particulier sur l'accès aux sources de données, aux data warehouse et datamarts…) il a une certaine parenté avec SPAD.

Utilisé dans 4 départements, en parallèle avec SPAD, il est apprécié des étudiants et enseignants et pourrait être appelé à voir son utilisation se généraliser. Certaines structures privilégiant désormais les logiciels libres (collectivités territoriales par exemple), il peut être intéressant d'initier les étudiants à un logiciel tel que Tanagra.

**KNIME**

La première version de Knime a été mise au point à l'université de Constance (Konstanz Information Miner) en 2006. Il s'agit d'un logiciel open source multiplateformes (Windows, Linux) d'apprentissage et de data-mining, téléchargeable gratuitement et d'installation très facile. Il comporte de nombreux modules de data management, d'analyse statistique et de graphiques. Il est assez complet : régressions linéaire, polynomiale, logistique, clustering, règles floues, arbres de décision, SVM, séries chronologiques…. Le concept de «pipeline » permet de décrire la chaîne de traitements : lecture des données, partitionnement de

l'échantillon, construction de la règle sur l'échantillon d'apprentissage, validation sur l'échantillon test…

Il a été utilisé par le département de Lisieux et a beaucoup intéressé les étudiants. L'expérience est limitée et récente, mais Knime semble apparaître comme une alternative aux autres logiciels de data-mining du fait de sa gratuité, du potentiel de ses possibilités dans la préparation des données, de leur analyse, modélisation et visualisation.

# 5 Une synthèse

## 5.1 Quels critères en STID pour le choix des logiciels ?

Lorsqu'on interroge les départements sur les raisons qui les ont conduits à faire tel ou tel choix de logiciel de traitement statistique, on voit apparaître :

- L'implantation du logiciel dans le monde professionnel. Il est impératif que les étudiants soient directement opérationnels sur les logiciels dominants du marché.
- Le coût du logiciel. Les départements doivent pouvoir acquérir les licences pour enseigner dans de bonnes conditions les logiciels choisis sans compromettre leur équilibre financier.
- Les qualités propres du logiciel : puissance, ergonomie, convivialité, possibilité de « tracer » une suite de traitements, souplesse en ce qui concerne le format des données en entrée, capacité à échanger avec d'autres logiciels…
- L'adéquation à une démarche pédagogique, à des étapes du processus d'apprentissage de la statistique. Au début du processus d'apprentissage, il peut être intéressant de faire de la statistique avec un logiciel tel que Excel très facile à utiliser et permettant d'entrer dans le détail des procédures de manière très fine. En revanche, en fin de cursus, les fondamentaux de la statistique étant mis en place, le travail sur l'acquisition du savoir-faire et la mise en œuvre des méthodes doit être exécuté avec des logiciels fournissant des procédures approfondies et rapides.
- La mise à disposition d'une version personnelle pour l'étudiant et l'enseignant. La possibilité pour l'étudiant de disposer d'une version personnelle durant ses études pour l'installer sur sa machine et travailler chez lui facilite beaucoup son travail. De même, que l'enseignant puisse disposer d'une version pour l'utiliser hors des salles d'enseignement ou de son bureau peut se révéler important.
- La possibilité pour l'étudiant de disposer du logiciel pendant son stage. L'étudiant peut être amené au cours de son stage à effectuer des analyses statistiques alors que l'entreprise ne dispose pas d'outil logiciel adapté pour le faire. Il est alors crucial que l'étudiant puisse disposer d'une manière ou d'une autre (prêt à titre gracieux, location de courte durée à tarif modique…) d'un logiciel adapté.

Actuellement on constate que SAS possède une position de premier plan dans la filière STID. Tous les départements enseignent SAS. Cela s'explique par son caractère exhaustif, sa robustesse, sa capacité à traiter des volumes très importants de données, et surtout sa position particulière dans le monde professionnel : SAS est incontournable actuellement dans le monde de la santé, mais est aussi fortement présent dans celui de l'industrie, des grandes

administrations, des assurances, mutuelles, banques, sociétés d'études marketing, sociétés de panel…

SPSS occupe une position notable en STID dans la famille des logiciels à menus déroulants. Il est d'utilisation facile (et donc peut être utilisé dès la 1re année) tout en offrant une très large palette de traitements statistiques. Mais SPSS ne possède pas une implantation dans le monde professionnel en France aussi forte que dans d'autres pays occidentaux.

Excel garde une position originale. Quasiment tous les départements ont recours à Excel à un moment du processus d'apprentissage de la statistique. Il est souvent considéré comme le logiciel de statistique pour non-statisticiens. Cette image vient de son utilisation par un public pour lequel la statistique se résume aux méthodes de statistique descriptive. Il n'en reste pas moins que certains départements ont recours à Excel même pour des traitements statistiques évolués.

Le logiciel R a pris en quelques années une place importante et cela pour plusieurs raisons : la gratuité, la possibilité de l'installer dans des environnements variés, la richesse de ses traitements statistiques. Il est cependant peu utilisé dans le monde professionnel. Les utilisateurs de logiciels du monde professionnel restent très attachés à des points tels que la compatibilité des versions successives, la garantie d'une exécution correcte, la présence d'un service « support clients », voire d'une hotline ; autant de points qui sont des freins au développement de R malgré ses qualités. On peut penser que R va progresser particulièrement dans les entreprises où le recours à la statistique est plutôt occasionnel et qu'il sera souvent présent dans les entreprises faisant de la statistique de manière régulière aux côtés des logiciels de statistique habituels.

En ce qui concerne les logiciels dédiés à des domaines particuliers de la statistique, MAPINFO pour les systèmes d'information géographique, SPAD pour le Data Mining et l'Analyse de Données, LE SPHINX pour le traitement des enquêtes, semblent occuper des positions stables, même s'ils cohabitent avec des logiciels gratuits.

## 5.2 La communauté STID et les éditeurs de logiciels

Les départements sont unanimes quant à leur insatisfaction à propos de leurs relations avec les éditeurs/distributeurs de logiciels :

- Les licences sont globalement jugées trop chères pour le budget des départements. Certains départements renoncent actuellement à remplacer d'anciens logiciels par de nouveaux, ou n'effectuent pas les mises à jour payantes, ou finissent par suspendre une location pour des raisons financières. Le sentiment général est que les éditeurs ne font pas l'effort qui devrait être fait en direction du monde de l'éducation. Il ne devrait pas être perdu de vue que nos diplômés pourront être amenés par la suite à guider leur employeur dans le choix d'un logiciel.
- Il existe de grandes disparités de traitements dues au fait que les négociations se font département par département. Les tentatives de la communauté STID pour obtenir des accords de tarification privilégiés et uniformes pour l'ensemble de la filière n'ont jamais reçu de réponse de la part des éditeurs/distributeurs. Il existe pour SAS un contrat privilégié dit « ministère » valable pour tout l'enseignement supérieur, mais il est limité à la partie basique du logiciel et ne concerne pas des modules tels que celui, par exemple, du Data Mining. D'autre part SAS est assez réticent à ce que les

départements aient recours à ce contrat et fait plutôt la promotion d'un système d'offres (diplôme ; université ; campus) généralement peu intéressant pour les départements.

– Les versions françaises de certains logiciels, c'est le cas par exemple pour Excel et pour SAS, sont souvent des traductions plus qu'approximatives de la version anglaise. Pour prendre un simple petit exemple parmi d'autres, la fonction LOI.NORMALE d'Excel utilisée sous la forme LOI.NORMALE(x ;0 ;1 ;1) renvoie la valeur au point x de la densité de la gaussienne centrée réduite. Or voilà ce qui est dit dans l'aide d'Excel : « la fonction renvoie la probabilité suivant une loi normale qu'un événement se reproduise x fois exactement » ! Ici la traduction n'est pas approximative, elle est tout simplement fausse. Lorsque les notions de fonction de répartition et de fonction de densité ne sont pas encore acquises de manière très stable par l'étudiant, il est certainement assez déroutant pour l'étudiant de tomber sur une telle phrase… Si ces problèmes de traduction ne sont pas trop graves quand l'utilisateur est déjà un statisticien expérimenté, il n'en va pas de même lorsqu'il s'agit d'un apprenant et il serait bon que l'éditeur écoute les utilisateurs sur ce sujet.

## 5.3 L'évolution à court terme de l'utilisation des logiciels dans la spécialité STID

Actuellement les départements ne souhaitent pas devoir réduire l'éventail des logiciels enseignés et espèrent que leur situation ne les conduira pas à le faire. Ils souhaitent poursuivre une pédagogie élaborée peu à peu au cours des dernières années : utiliser des logiciels adaptés pour l'acquisition des bases de la statistique, former les étudiants aux grands logiciels utilisés dans le monde professionnel et disposer aussi des logiciels dédiés, quand c'est nécessaire, à certains domaines particuliers de la statistique.

Cependant le coût des licences est lourdement ressenti dans une période de stagnation ou de diminution des budgets de fonctionnement. Le risque est important de voir les départements se tourner de plus en plus vers des logiciels gratuits, avec les conséquences qui en découlent pour le caractère professionnalisant de la formation. Au cours de la rédaction de cet article, l'éditeur StatSoft du logiciel STATISTICA a proposé un partenariat aux départements STID qui va permettre à ceux-ci de l'enseigner à partir de l'année universitaire 2012-2013.

On parle maintenant du Cloud Computing et en particulier de la solution SaaS (Sofware as a Service) comme d'une évolution très probable de la façon de travailler avec les logiciels. Dans le cadre d'une telle solution, les départements ne disposeraient plus des logiciels en local (i.e. sur les postes des étudiants ou sur serveur local), mais s'abonneraient à un service leur permettant d'utiliser, via internet, les logiciels situés sur un serveur distant. Même si cette solution commence à connaître un développement sensible dans certains secteurs professionnels, il est encore tôt pour savoir si elle va se développer dans le monde de l'éducation, à quel prix et quelles seraient les conséquences pour nos départements de la mise en place de ce mode de travail.

## Références

[1] Programme pédagogique national des départements STID : http://media.enseignementsup-recherche.gouv.fr/file/DUT_-_Programmes_pedagogiques_nationaux/83/2/statistique_et_informatique_decisionnelle_157832.pdf

[2] STID-France : http://www.stid-france.com/ http://fr-fr.facebook.com/pages/STID-FRANCE/178991655484834

[3] EXCEL : http://office.microsoft.com/fr-fr/

[4] SAS Institute : http://www.sas.com

[5] R : http://cran.r-project.org/

[6] SPSS : http://www-01.ibm.com/software/fr/analytics/spss/

[7] SPAD : http://spad.eu/spad/index.php?categ=accueil&page=presentation

[8] MAPINFO : http://www.pbinsight.com/welcome/mapinfo/

[9] XLSTAT : http://www.xlstat.com/fr/

[10] STATGRAPHICS : http://www.statgraphics.fr/

[11] MINITAB : http://www.minitab.com/fr-FR/

[12] LE SPHINX : www.lesphinx-developpement.fr

[13] ARCGIS : http://www.esrifrance.fr/arcgis.asp

[14] LIMESURVEY: http://www.limesurvey.org/fr

[15] TANAGRA : eric.univ-lyon2.fr/~ricco/tanagra/

[16] KNIME: http://www.knime.org/

[17] STATISTICA : http://www.statsoft.fr/index.php