

# L'ENSEIGNEMENT À DISTANCE EN STATISTIQUE ET L'USAGE DE LOGICIELS AU TRAVERS DE DEUX EXEMPLES

Marthe-Aline JUTAND<sup>1</sup> et Anne RUIZ-GAZEN<sup>2</sup>

## TITLE

E-learning Statistics and software use: two examples

## RÉSUMÉ

L'offre de formation via Internet en statistique se développe depuis une dizaine d'années à visée de profils d'apprenants très différents. La mise en œuvre de ces formations implique une réflexion toute particulière de par la non proximité des étudiants et parfois une disponibilité non régulière de ces derniers. Il s'avère donc nécessaire de réfléchir spécifiquement à la formation, et la question du choix de logiciels dans le cadre de ces apprentissages est un élément important. Les expériences d'enseignement de statistique via Internet de l'ISPED Université Bordeaux Segalen et de l'Université de Toulouse 1 Capitole sont des sources de réflexion à ce sujet.

*Mots-clés* : enseignement à distance, enseignement de la statistique, pédagogie.

## ABSTRACT

Since ten years, the e-learning in statistics has been developed not only for statisticians but for very different learners' profiles. The implementation of these programs requires particular thinking because the students are isolated and generally not fully available. It is therefore necessary to consider the choice of software to obtain the best result. The experiences of the ISPED-University Bordeaux Segalen formations (University diploma and public health master) and the University Toulouse 1 Capitole are presented as possible approaches.

*Keywords*: e-learning, pedagogy, statistics teaching.

## 1 Introduction

L'enseignement à distance existe depuis déjà de nombreuses années sous différents formats. Il s'agissait tout d'abord de formations faisant appel à un échange postal régulier de documentation et résolution d'exercices ou de devoirs permettant ainsi d'évaluer l'avancée de l'apprentissage. C'est par exemple le cas du CNED (Centre National d'Enseignement à Distance) en France. Depuis une quinzaine d'années, l'usage d'Internet s'étant démocratisé, l'enseignement a pu développer une offre de formation ouverte à distance. Une Formation Ouverte À Distance (FOAD) est une action de formation souple reposant sur des situations d'apprentissage complémentaires et plurielles en termes de temps, de lieux et d'actions. Elle n'est pas nécessairement exécutée sous le contrôle permanent d'un formateur et elle peut se dérouler partiellement ou totalement à distance<sup>3</sup>.

---

<sup>1</sup> ISPED, Université Bordeaux Segalen, marthe-aline.jutand@isped.u-bordeaux2.fr

<sup>2</sup> Université Toulouse 1 Capitole, anne.ruiz-gazen@tse-fr.eu

<sup>3</sup> Financement et mise en œuvre de la FOAD : vade-mecum des bonnes pratiques, octobre 2011, <http://www.fffod.org/media/20111004-vademecum.pdf>

Il s'agit dans cet article non pas de faire une revue des plate-formes d'e-learning de statistique comme cela est présenté dans d'autres articles (Härdle *et al.*, 2006), mais de présenter via l'expérience de l'Institut de santé publique, d'épidémiologie et de développement de l'Université Bordeaux Segalen et celle de l'Université de Toulouse 1 Capitole, les atouts, les intérêts mais aussi les difficultés de l'apprentissage de la statistique via Internet, et d'insister tout particulièrement sur la mise en œuvre de l'apprentissage de logiciels. Ces deux expériences, bien que différentes au premier abord de par les publics ciblés, se trouvent en réalité confrontées aux mêmes réalités et questionnements pouvant ressembler à ceux de nombreuses autres expériences.

## 2 Présentation des deux situations

### 2.1 Développement historique de ces offres de formation

#### 2.1.1 Au sein de l'ISPED - Bordeaux

L'Institut de santé publique, d'épidémiologie et de développement (ISPED) est une composante de l'Université de Bordeaux Segalen. Sa mission principale de formation est d'apporter aux étudiants toutes les compétences qui leur permettront de contribuer à relever les grands défis posés par la santé publique contemporaine : l'accroissement de l'espérance de vie, la refonte des systèmes de santé, la résurgence des maladies infectieuses dans le monde, l'impact de l'industrialisation et de la mondialisation sur l'environnement et la santé des populations. La forte association de l'ISPED au Centre de Recherche INSERM U897 Epidémiologie et Biostatistique offre un cadre privilégié pour son offre de formation qui se décline en un ensemble de diplômes d'université (environ 25) et un master de santé publique dont la première année correspond à un tronc commun.

L'offre de formation de DU s'est particulièrement développée depuis 2001 (<http://ead.isped.u-bordeaux2.fr>) par le biais de la formation via Internet et, tout particulièrement, en réponse à une demande de formation francophone en santé publique d'un public ne pouvant pas accéder à des formations en présentiel. A ce jour, 11 DU sont proposés via Internet, parmi lesquels deux sont à dominante statistique (DU méthodes statistiques en santé et DU méthodes statistiques de régression en épidémiologie).

Les différents enseignements des DU se retrouvant, pour la plupart, dans le contenu pédagogique du master de santé publique, il a été décidé dès 2007 de proposer une modalité d'enseignement via Internet pour la première année de master et la seconde année spécialité épidémiologie. Ainsi, des étudiants ayant validé certains DU pouvaient compléter ces formations pour obtenir un master. L'enseignement du master proposé en ligne correspond exactement au contenu d'enseignement proposé en présentiel à quelques options près. A ce jour, le nombre d'inscrits est de 90 en DU méthodes statistiques en santé, 40 en DU méthodes statistiques de régression en épidémiologie, 34 en première année de master et 14 en seconde année de master spécialité épidémiologie.

#### 2.1.2 A l'Université Toulouse 1 Capitole

En 2000, l'Université Toulouse 1 a choisi de développer l'enseignement à distance (<http://foad.univ-tlse1.fr>). Depuis septembre 2003, l'équipe toulousaine d'enseignants du master 2 statistique et économétrie propose une version entièrement à distance et diplômante

*M.-A. Jutand et A. Ruiz-Gazen*

des cours offerts en présentiel. Depuis la rentrée 2009, l'Université Toulouse 1 Capitole a aussi ouvert une année préparatoire au master 2 ainsi qu'un diplôme d'université en « statistique appliquée ».

La version entièrement à distance du master statistique et économétrie est née de la conjonction de différents événements. Des demandes de perfectionnement de la part de professionnels arrivaient régulièrement mais elles ne pouvaient être que difficilement satisfaites par la version présentielle du master, lourde, avec un emploi du temps ne permettant pas de planifier la participation aux cours plusieurs mois à l'avance, et supposant que l'étudiant réside à Toulouse.

Un des points forts de la formation est de s'appuyer sur une compétence solide à la fois dans le domaine statistique et dans le domaine économique, grâce à son double ancrage dans le Groupe de Recherche en Economie Mathématique et Quantitative (GREMAQ) et l'Institut de Mathématiques de Toulouse (IMT), unités mixtes de recherche CNRS, mais aussi grâce à l'intervention de professionnels. La formation est fortement ancrée sur les besoins effectifs des entreprises et les étudiants sont tout de suite opérationnels après le master.

Comme le master 2 se déroule sur 2 ans avec un recrutement annuel et qu'il est complété par une année préparatoire et un DU, une soixantaine d'étudiants sont actuellement inscrits dans les filières à distance en statistique à UT1-Capitole. La plupart des étudiants sont européens ou africains.

### **2.1.3 Points communs**

Les formations présentées ci-dessus sont soutenues par l'Agence Universitaire de la Francophonie (AUF) et proposent actuellement une offre de formation graduelle, partant de DU pour arriver à l'obtention d'un master. Cette offre de formation permet ainsi de contenter un large panel de candidats, allant de ceux qui souhaitent juste obtenir une mise à jour de leurs connaissances ou un apprentissage, en suivant un DU, à ceux qui souhaitent pouvoir obtenir un niveau master qui va apporter une importante plus-value dans leur parcours professionnel futur.

## **3 Organisation pédagogique**

La création de formations numériques nécessite le développement de nouvelles approches pédagogiques qui doivent tenir compte du profil et contexte des apprenants. Ces offres de formations sont souvent très attractives pour un public de professionnels et donc d'adultes souhaitant compléter leur formation tout en travaillant.

L'ingénierie pédagogique et de formation mise en œuvre dans le cadre de l'élaboration de ces formations via Internet peut se décrire selon les cinq phases proposées dans le modèle ADDIE, c'est-à-dire : Analyse, Design, Développement, Implémentation et Evaluation (Molanda *et al.*, 1996). C'est selon ce modèle que nous allons détailler la mise en place de nos formations, les choix faits pour ce qui concerne les moyens et outils pertinents et tout particulièrement pour ce qui est de l'apprentissage et la mise en œuvre de logiciels.

### 3.1 Phase d'analyse : différents types de publics

La formation via Internet s'est développée au sein de nos deux établissements suite à une demande de plus en plus importante de formations délocalisées (principalement en direction des professionnels d'Afrique sub-saharienne) due aux difficultés rencontrées par les professionnels étrangers pour venir se former en France. La demande croissante des professionnels français souhaitant avoir accès à la formation continue sans devoir quitter leur emploi a aussi été un indicateur mettant en évidence la nécessité de proposer ce type de formation via Internet.

#### 3.1.1 Santé publique - Bordeaux

De par le domaine de compétence de l'ISPED, la demande était en particulier issue de médecins ou autres professionnels du champ de la santé publique. L'essentiel de ce public n'a jamais eu d'enseignement spécifique en statistique et éprouve, pour la plupart, la nécessité d'aborder ce champ de connaissance tant les données chiffrées et les résultats d'études sont déterminants dans les choix de politique de santé publique. Il s'agit donc de leur permettre d'acquérir des compétences non pas de statisticiens, mais de professionnels ayant tous les jours besoin d'avoir une lecture critique d'un certain nombre d'informations communiquées. De plus, ils doivent être en mesure de mettre en œuvre une analyse descriptive de fichiers de données, mais aussi d'utiliser des tests statistiques classiques ainsi que des méthodes de régression linéaire et logistique et des modèles de survie peu sophistiqués. Cela leur permettra ainsi d'être en mesure de travailler en étroite collaboration avec des statisticiens.

Cette étape d'analyse a permis de souligner l'importance de mettre en place un outil d'enseignement souple de par les contraintes logistiques importantes, ainsi qu'un accompagnement de proximité.

Il s'agissait de plus de donner des outils permettant aux apprenants de mettre en œuvre les savoirs et ainsi d'équilibrer les parties entre théorie et application. Pour cela il était important de définir un choix pertinent de logiciels statistiques leur donnant la possibilité de mettre en œuvre les méthodes apprises sans se trouver face à des difficultés insurmontables de par la non présence d'un enseignant. Les critères de choix pour les logiciels étaient donc : accessibilité quel que soit le lieu de connexion, pertinence dans le domaine d'application, et applicabilité de l'ensemble des méthodes vues dans le cadre de l'enseignement.

Afin de lutter contre l'isolement de l'étudiant, des séances de regroupement ont été décidées durant le master, deux fois par an.

#### 3.1.2 Statistique et Econométrie - Toulouse

A la différence des formations proposées à l'ISPED, les formations de Toulouse, et tout particulièrement le master, accueillent des étudiants pourvus d'une solide culture statistique, académique ou empirique. En formation initiale, au niveau master 2, le recrutement se fait parmi les titulaires d'un diplôme de niveau Bac+4, obtenu dans d'excellentes conditions, en économétrie ou en statistique mathématique. En formation continue ou reprise d'études, il intéresse des statisticiens d'entreprises et des secteurs public et para public ayant une bonne formation de base en statistique, validée par un diplôme de niveau Bac+3/4, la pratiquant régulièrement et désireux d'améliorer leur qualification. Il intéresse également des ingénieurs

d'autres disciplines appelés à utiliser l'outil statistique, des statisticiens confirmés qui veulent rafraîchir ou mettre à jour leurs connaissances sur un point particulier.

L'année préparatoire et le DU de « Statistique Appliquée » s'adressent davantage à des professionnels qui ont besoin d'une remise à niveau sur les méthodes mathématiques utiles en master 2 mais aussi sur les méthodes de base en analyse de données exploratoire et en économétrie.

Le master 2 Statistique et économétrie en FOAD a pour objectif de former des chargés d'études spécialistes en statistique et économétrie. Il s'agit donc de perfectionner des connaissances et des compétences déjà acquises au cours de formations précédentes.

Les étudiants ont donc en général déjà eu un apprentissage de certains logiciels statistiques mais il est finalement assez rare qu'ils connaissent les logiciels qui sont proposés dans la formation à distance, à savoir SAS et R. Ces deux logiciels ont été choisis pour différentes raisons. La raison principale est qu'il s'agit de logiciels qui permettent d'avoir accès à de très nombreux et de très sophistiqués traitements statistiques. Grâce à ces deux logiciels, les étudiants peuvent mettre en œuvre n'importe laquelle des méthodes étudiées dans les différents cours. Le grand avantage de R est bien sûr le fait qu'il s'agit d'un logiciel libre très facilement téléchargeable et donc utilisable n'importe où, même dans de très petites entreprises. Concernant SAS, le prix et la difficulté du téléchargement ne sont clairement pas des atouts, mais la puissance du logiciel dans la gestion des bases de données et son importante utilisation dans de nombreuses grandes entreprises sont un réel plus pour le CV des étudiants de nos formations.

### **3.1.3 Points communs**

Pour Bordeaux et Toulouse, d'un point de vue de l'organisation et après la phase d'analyse, il a été décidé que, de par les contraintes de lieu et de temps, chaque cours se déroule de façon asynchrone sur un trimestre. Le choix a donc été fait d'une formation complètement à distance, à l'exception pour le master de santé publique de deux périodes de regroupement par an. De plus, l'étudiant est accompagné tout au long de sa formation par le biais de forums ou d'échanges par courriels.

Pour ne pas laisser l'étudiant seul face à son organisation, pour chaque formation un calendrier annuel est diffusé à l'étudiant, et la remise de travaux rythme l'année scolaire. Presque tous les modules d'enseignement en statistique sont évalués par un projet à rendre qui nécessite l'utilisation d'un logiciel statistique, pour la mise en application.

## **3.2 Phase de design : différents types de publics**

### **3.2.1 Santé publique - Bordeaux**

Après la phase d'analyse il est apparu qu'il y avait différents profils parmi les apprenants. Les objectifs pédagogiques en statistique ont été cependant assez facilement définis avec les enseignants d'épidémiologie car il s'agissait tout particulièrement d'être complémentaire afin de développer les compétences nécessaires pour un futur professionnel en santé publique.

Certains apprenants n'avaient pas de connaissance en statistique et il était donc nécessaire de reprendre la base, sans s'éterniser sur les développements mathématiques mais en développant plutôt l'approche par problèmes concrets de leur domaine d'application. Le DU

*L'enseignement à distance en statistique et l'usage de logiciels au travers de deux exemples*

« méthodes statistiques en santé » répond à ces attentes. Il est composé de 4 modules : introduction et statistique descriptive, fluctuations d'échantillonnage et estimation, tests statistiques et analyse de variance, et régression linéaire simple.

Pour d'autres apprenants il s'agissait réellement d'approfondissement spécifique des méthodes de régression très utilisées dans le champ de la santé publique et tout spécifiquement en épidémiologie. Pour cette raison un second DU a été proposé ; il s'agit du DU « méthodes statistiques de régression » composé aussi de 4 modules : introduction à l'analyse multivariable et aux modèles de régression, méthodes de régression linéaire simple et multiple, régression logistique et méthodes d'analyse de données de survie.

Le contenu pédagogique de ces deux DU couvre l'ensemble du programme de statistique de la première année de master de santé publique, tandis que le programme de statistique de seconde année de la spécialité épidémiologie équivalent à 60 heures d'enseignement complète la formation par une introduction aux modèles linéaires généralisés et à la régression de Poisson, aux méthodes d'analyse des données de survie et modèle de Cox, à la gestion des données manquantes et imputation multiple, à l'analyse des données longitudinales et enfin à l'analyse des données multidimensionnelles.

Les études de cas font appel bien évidemment à l'usage de logiciels permettant la mise en œuvre des outils statistiques. Il était donc nécessaire de donner les moyens aux étudiants de manipuler des logiciels et, pour ce faire, de choisir des logiciels, de proposer des tutoriels et de développer les exemples de cours avec les différents logiciels.

Initialement, les logiciels que l'on souhaitait utiliser étaient Excel, EpiInfo et SAS tel que pour les enseignements en présentiel, mais rapidement le logiciel R a été rajouté à cette liste pour des raisons qui seront développées ci-dessous.

Le logiciel SAS a été proposé uniquement pour les étudiants inscrits en master de santé publique, de par la difficulté à faire parvenir ce logiciel à l'ensemble des inscrits en DU mais aussi la difficulté de son apprentissage pour un étudiant isolé. La gratuité de R quant à elle permet de faire la présentation de cet outil à l'ensemble des étudiants sans en rendre obligatoire son usage pour les DU.

Le logiciel EpiInfo, quant à lui, est un logiciel de traitement de données conçu pour des épidémiologistes. Complet, il permet de mettre en œuvre tout ce qui concerne la partie informatizable du traitement des données issues d'une enquête. Cela comprend notamment : la mise en forme du formulaire de saisie, la saisie et l'analyse statistique des données. Très ciblé, il permet d'obtenir rapidement des résultats utiles. EpiInfo est un logiciel gratuit, développé par les CDC (Centers for Disease Control) à Atlanta et librement diffusé. Les dernières versions sont téléchargeables à partir du site des CDC : <http://www.cdc.gov/epiinfo/>. Ce logiciel couvre l'analyse descriptive, les tests d'hypothèses basiques (comparaison de moyennes, Mann-Whitney, test de chi deux d'indépendance, test de Fisher), régression linéaire, régression logistique. Ce logiciel n'offre pas une large amplitude de tests statistiques et est aussi limité en outils graphiques ; cependant sa gratuité, sa facilité d'utilisation et le fait que l'on puisse développer un questionnaire, saisir les données et les analyser en font un outil très apprécié et pratique. L'usage de cet outil est, de plus, fréquent dans le cadre d'ONG dont sont issus un certain nombre de nos étudiants suivant les formations à distance.

### 3.2.2 Statistique - Econométrie Toulouse

Dès le début, en 2003, il a été décidé que le contenu du master 2 FOAD serait équivalent à celui de la formation en présentiel. La formation de master 2 FOAD a été répartie sur deux ans et non sur une seule année comme en présentiel, l'évaluation du temps d'apprentissage étant de 840h. Les douze cours que l'étudiant doit suivre sont donc répartis entre les deux années.

Ces douze cours couvrent une grande partie de la Statistique et sont d'un niveau avancé. Certains cours sont des cours de statistique plutôt généralistes comme les cours sur les modèles linéaires gaussiens, le data mining (méthodes d'apprentissage), les sondages, les séries temporelles et les modèles non-paramétriques. Les cours de logiciels statistiques et de bases de données sont plutôt des cours d'informatique. Trois cours sont sur des domaines d'applications plus précis qui intéressent tout particulièrement le public que nous formons : les cours d'économétrie, de scoring et de statistique de l'assurance. Enfin, les étudiants ont le choix entre deux options : ou bien un cours d'économétrie spatiale et un autre de statistique spatiale ou bien un cours de web mining et un cours de durée de vie.

Cette maquette est très proche de la maquette actuelle du master 2 statistique et économétrie en présentiel, la principale différence résidant dans le nombre moindre d'options en FOAD comparativement au présentiel.

Concernant le DU en « Statistique appliquée », il a pour objectif d'approfondir des connaissances en statistique au niveau BAC+4 et d'acquérir des compétences dans l'utilisation des logiciels statistiques SAS (facultatif) et R. Le DU est constitué de 3 matières : logiciels statistiques, data mining (exploration) et économétrie. L'année préparatoire au master 2 statistique et économétrie consiste en les matières du DU complétées par deux cours de remise à niveau en probabilités, algèbre linéaire et optimisation qui sont nécessaires pour suivre le master 2 dans de bonnes conditions, surtout pour des étudiants qui n'ont pas fait de mathématiques depuis plusieurs années.

Presque tous les cours, que ce soit en DU ou en master 2, sont assortis d'exercices et de projets sur de vrais jeux de données mis en œuvre avec des logiciels statistiques. Cela signifie que les étudiants ne se voient pas uniquement proposer la théorie statistique mais qu'ils ont la possibilité de mettre en pratique toutes les analyses. L'accent est aussi mis sur les interprétations des résultats obtenus en sortie des logiciels statistiques. Ainsi, le cours de logiciels statistiques qui est enseigné dès le premier trimestre en DU comme en master 2 est un cours fondamental pour les apprenants puisqu'il permet aux étudiants de se familiariser avec les logiciels qu'ils vont utiliser tout au long de leur formation.

Au moment du lancement de la formation, les principaux logiciels statistiques utilisés dans le master 2 étaient Matlab et SAS. Rapidement, certains enseignants ont proposé l'utilisation de R qui s'est ensuite généralisé, Matlab n'étant plus utilisé actuellement dans nos formations. Les difficultés dans le téléchargement de SAS nous ont conduits récemment à demander aux enseignants d'offrir aux étudiants la possibilité de mettre en œuvre les analyses étudiées en R (obligatoire) et éventuellement en SAS mais à ne plus exiger SAS pour les examens puisque certains étudiants n'y ont pas accès. Pour faciliter le passage au logiciel R de certains enseignants qui ne travaillaient qu'en SAS, nous avons organisé au Gremaq une journée de formation à R pour les enseignants-chercheurs avec une partie initiation au logiciel et une partie plus avancée.

### 3.2.3 Points communs

Dans les deux cas la stratégie pédagogique mise en œuvre était donc de proposer des contenus appropriés aux attentes et besoins des candidats et faisant appel à des exemples concrets de leur domaine. Il ne s'agissait pas de proposer une énième formation en statistique.

Les méthodologies retenues ont été un enseignement asynchrone associant toujours la diffusion de contenu pédagogique via des supports en traitement de texte mais aussi avec des animations en richmédia par exemple et s'appuyant toujours sur des exemples ainsi que la mise en pratique des notions abordées avec manipulation de fichiers de données.

Le choix des logiciels à utiliser était évidemment primordial. Les critères de gratuité des logiciels et d'accessibilité sont deux critères très importants lorsqu'on souhaite travailler avec des apprenants pouvant venir de régions et pays très différents, pouvant être plus ou moins privilégiés. Ces critères expliquent par exemple le choix de développer l'usage de R pour les deux formations et d'EpiInfo pour la formation en santé publique.

Le choix du logiciel Excel de la suite Office de Microsoft s'explique par son accessibilité et la connaissance qu'en ont la plupart des apprenants inscrits à ce type de formation à distance.

Le logiciel SAS, de par l'importance de sa part de marché dans le domaine de la statistique, est ou était un logiciel inévitable, tout particulièrement pour les étudiants du nord. Mais le développement et la gratuité de R a amené les équipes pédagogiques à privilégier cet outil et à développer les supports pédagogiques appropriés.

Dans les deux cas, lorsqu'il fut décidé de se lancer dans l'aventure de la formation à distance, la mise en marche fut assez douloureuse. Les auteurs dont les supports de cours se limitaient à des diapositives, complétées par de généreux commentaires oraux, ont dû étoffer leurs présentations pour les rendre étudiables de façon autonome. La réalisation de supports de cours adaptés a nécessité du temps. Un travail important a dû aussi être réalisé pour accompagner les apprenants dans la prise en main des différents logiciels proposés. Tout cela constitue le temps de la phase de développement.

## 3.3 Phase de développement : plate-forme d'enseignement et usage de logiciels

### 3.3.1 Santé publique - Bordeaux

Depuis 2001, la plate-forme d'enseignement à distance (EAD) PLEI@DE (Programme en Ligne d'Enseignement Interactif À Distance) est développée par le Centre de Recherche et de Développement en Informatique Médicale (CREDIM-Université Bordeaux Segalen).

Cette plate-forme d'enseignement va permettre à chaque étudiant inscrit d'avoir accès aux enseignements sous différents formats, cours en pdf ou page internet comportant ou non des animations. Il trouvera aussi des exercices interactifs, des tests de validation, des fichiers pour les études de cas, mais aussi l'ensemble des informations administratives dont il a besoin pendant sa formation. Les forums, le chat, ainsi qu'une fonctionnalité de partage de documents sont autant d'outils qui permettront à l'étudiant d'avoir des échanges avec les enseignants mais aussi avec les autres apprenants.

Une équipe informatique développe régulièrement de nouveaux outils suivant les demandes d'enseignants, des services administratifs mais aussi suite aux évaluations complétées par les étudiants à l'issue de la formation. Il est très important de souligner que la mise en ligne des contenus pédagogiques n'est pas effectuée directement par l'enseignant mais par une équipe de deux personnes que nous appelons « médiatiseuses ». Leur rôle est central car elles suivent la mise en ligne des documents d'enseignement pour respecter les calendriers, elles suggèrent des améliorations dans la présentation des contenus. Elles sont un élément clé pour la formation des nouveaux enseignants dans la prise en main de la plateforme. Il est très important de souligner que cette organisation basée sur un service support pour les enseignants est certainement une des clés de la réussite de cette offre de formation.

Cependant, avant de diffuser les outils pédagogiques, il est bien nécessaire de les créer. Pour leur développement, et afin que les étudiants puissent appréhender leur apprentissage comme un cumul de connaissances, il était nécessaire que les enseignants travaillent en étroite collaboration. Ainsi chaque enseignant était tour à tour rédacteur ou lecteur de contenu pédagogique. Les choix et les notations d'écriture scientifique ont été standardisés pour l'ensemble de l'enseignement actuel et à venir. Ce travail de construction collective d'outils et supports pédagogiques a été particulièrement formateur et très favorisant aussi pour la formation en présentiel. La construction de l'enseignement a été pensée aussi autour d'exemples issus d'enquêtes liées à la santé, qui servent ainsi de fils conducteurs tout au long des cours et permettent d'illustrer les différentes notions abordées.

Dans cette phase de développement de l'enseignement de la statistique, il était très important de développer des tutoriels permettant d'accompagner l'étudiant dans la prise en main des logiciels qui lui étaient indiqués. Un travail important a donc été conduit pour proposer aux étudiants des documents téléchargeables d'aide mais aussi un certain nombre d'animations sous le logiciel Adobe Captivate, permettant de gérer des contenus de e-learning. De plus, des exercices avec correction d'utilisation du logiciel sont proposés avant de développer les parties propres au contenu pédagogique de l'enseignement de statistique. Un exemple de développement de contenu pédagogique pour le logiciel EpiInfo est accessible à partir du campus numérique de l'ISPED<sup>4</sup> sous l'intitulé « [essentiel de santé publique](#) ».

Des tutoriels ont donc été développés pour le logiciel EpiInfo, SAS, R et quelques éléments pour Excel (graphiques et tableaux dynamiques). Ces supports sont d'ailleurs maintenant largement utilisés pour les formations en présentiel.

### **3.3.2 Statistique - Econométrie Toulouse**

Toulouse 1 Capitole a choisi comme plate-forme d'apprentissage en ligne pour l'ensemble de ses formations ouvertes à distance la plateforme Moodle qui est sous licence libre. Les échanges entre enseignants et apprenants sont facilités car cette plate-forme offre des fonctions pédagogiques d'apprentissage en ligne efficaces, simples d'utilisation et conviviales. De plus, l'équipe des techniciens du service FOAD d'UT1 Capitole ainsi que l'équipe TICEA (Technologies de l'Information et de la Communication pour l'Enseignement et l'Apprentissage) offrent aux enseignants différentes formations à la plate-forme. L'équipe FOAD offre des formations individuelles si besoin est et une assistance quotidienne pour tous les problèmes que peuvent rencontrer les enseignants ou les étudiants de la FOAD avec la plate-forme. L'équipe TICEA propose quant à elle des formations avancées pour aider à la

---

<sup>4</sup> <http://ead.isped.u-bordeaux2.fr/>

*L'enseignement à distance en statistique et l'usage de logiciels au travers de deux exemples*

mise en place d'enseignement via Moodle. Elle se tient à jour aussi des dernières avancées dans le domaine de l'enseignement via Moodle et les répercute aux enseignants. Ainsi, nous sommes en train de bénéficier de la mise en place de logiciels anti-plagiat utilisables facilement depuis Moodle.

L'utilisation de la plate-forme varie sensiblement selon les enseignants car nous n'avons pas souhaité imposer une façon de procéder. Toutefois, au moment du lancement du projet l'équipe pédagogique a beaucoup discuté et collaboré sur la façon d'enseigner via la plate-forme et nous avons ensuite évolué ensemble en discutant au moins une fois par an, lors de la réunion de rentrée ou sur l'espace de cours Moodle réservé aux seuls enseignants, de nos diverses expériences, réussies ou pas. La rédaction des supports de cours et des exercices et travaux pratiques sur ordinateur avec correction est particulièrement lourde au départ pour les enseignants (rétribués en droits d'auteurs pour cette tâche) et suppose que l'équipe enseignante soit motivée par ce type d'enseignement. Les cours sont généralement au format pdf et sont écrits par les enseignants sans assistance particulière.

Généralement, nous fournissons aux étudiants des supports de cours mais aussi des exercices corrigés et des exercices à chercher au fur et à mesure du trimestre. Ces exercices sont théoriques ou pratiques et nous fournissons des corrections détaillées. Nous demandons ensuite aux étudiants de réaliser un projet sur des données réelles avec les logiciels statistiques R ou SAS. Pour plus de détails sur le fonctionnement et un exemple de pédagogie mise en œuvre avec la plate-forme de télé-enseignement Moodle, nous renvoyons le lecteur à Aragon *et al.* (2007). L'outil de communication privilégié entre étudiants et enseignants est le forum et il en existe en général plusieurs par cours. L'importance des forums comme lieu de partage des interrogations des apprenants, de centralisation des réponses des enseignants, de discussion des cours, ne nous était pas manifeste au début de notre expérience mais est apparu clairement après un ou deux ans. L'utilisation du chat et des vidéos est plus rare dans nos cours mais elle existe toutefois. Ainsi, le cours de logiciels statistiques qui a lieu au début de nos formations contient plusieurs vidéos. Concernant le chat, il est utilisé notamment dans le cours de sondages (deux ou trois séances) car le contact direct est particulièrement apprécié de certains étudiants. Enfin, quelques cours contiennent des activités Quiz qui permettent aux étudiants de réviser le cours avant de commencer les exercices.

Le master 2 statistique et économétrie de Toulouse est une formation de haut niveau en statistique ; par conséquent, les intervenants sont, d'une part, des professionnels au fait des meilleures pratiques de leur métier mais également très pris par leur profession, d'autre part, des enseignants-chercheurs qui, à côté de l'enseignement, poursuivent une carrière de chercheur dans un domaine proche de ce qu'ils enseignent. Il est donc important de préciser les méthodes de travail communes en définissant clairement les contraintes en particulier au niveau de l'emploi du temps qui ne doit pas déborder du calendrier prévu. Il est à noter aussi que l'équipe d'enseignants dans nos formations à distance a beaucoup évolué (départs à la retraite oblige) depuis les débuts du master 2 FOAD. Les enseignants qui nous ont rejoints ces dernières années n'avaient pas en général d'expérience de l'enseignement à distance et ils ont été informés en détail de la façon dont nous procédons. Ils ont aussi accès à d'autres espaces de cours pour pouvoir comprendre facilement le fonctionnement à partir d'exemples.

### 3.3.3 Point commun

Les personnes choisissant de suivre une formation via Internet ont déjà l'habitude de l'usage de l'ordinateur et d'un certain nombre d'outils dont Microsoft ou Open Office. Pour

cette raison ils ont pour la plupart utilisé Excel ou Open Office Calc et leurs pré-requis leur permettent facilement d'utiliser ces logiciels pour la partie statistique descriptive.

Le logiciel SAS quant à lui est très répandu dans le monde de la santé mais aussi dans le milieu des banques et des assurances, et la connaissance de SAS est souvent un indispensable pour postuler à certaines offres d'emplois. L'apprentissage de ce logiciel connu comme difficile pour les formations en présentiel, devient très complexe dans le cadre de formation à distance. Il est donc nécessaire de développer des documents supports très explicites décomposant l'ensemble des étapes de chaque analyse. Un tutoriel a donc été développé pour chaque formation et l'ensemble des exemples et des corrections d'exercices ont été réalisés.

Le logiciel R est maintenant enseigné dans la plupart des formations, sa gratuité en a fait un de ses atouts majeurs et tout particulièrement pour les formations à destination d'un public parfois délocalisé dans des pays à ressources limitées, comme celles présentées ici. Son téléchargement rapide est aussi un facteur facilitant. Cependant il s'agit d'un outil sophistiqué qui nécessite une aide par tutoriel pour accompagner sa prise en main.

Le développement des supports d'enseignement pour la formation en ligne nécessite la prise en considération de l'éloignement de l'étudiant et son isolement. Pour ces différentes raisons il est nécessaire de développer des outils appropriés à la discipline et le plus interactif possible pour que l'étudiant se sente accompagné comme s'il avait pu bénéficier d'une formation en présentiel. Il est difficile de restituer l'ambiance d'un TD en salle informatique dans le cadre d'une formation à distance, mais cependant il est important de développer des outils permettant de répondre aux questions de l'étudiant et surtout de ne pas le laisser perdu seul face à son écran en ne sachant pas quoi faire. Le développement de forum ou de chat sont des outils utiles mais cependant difficilement quantifiables en temps d'enseignement.

### **3.4 Phase d'implémentation**

Une fois que l'enseignement a été construit, que l'ensemble des outils ont été développés et que la plate-forme est en service, le moment crucial de l'ouverture de la formation est enfin arrivé. Dans un premier temps il est nécessaire de former les nouveaux étudiants à l'usage de la plate-forme<sup>5</sup>, ainsi qu'aux différents outils qu'ils doivent manipuler et donc tout particulièrement aux logiciels.

#### **3.4.1 Santé publique - Bordeaux**

Chacune des formations à distance, DU et master, est coordonnée par un responsable de formation, un responsable administratif et une équipe pédagogique. L'ensemble des enseignants en statistique dans l'une ou l'autre de ces formations se réunissent une fois par semestre pour faire un point de l'état d'avancement de la formation.

Afin de pouvoir être réactif aux problèmes rencontrés par les étudiants, il a été décidé qu'un forum serait ouvert pour chaque unité d'enseignement pour le master et chaque module pour les DU. Le responsable de l'enseignement a donc pour mission de répondre aux demandes des étudiants qui seront postées mais aussi de régulièrement les solliciter afin de ne pas les laisser seuls face à leurs interrogations. Les étudiants en formation à distance peuvent rencontrer en réalité les mêmes difficultés que les étudiants en présentiel lorsqu'il s'agit de poser des questions en public. On pourrait imaginer que les étudiants se sentent moins

---

<sup>5</sup> <http://campus.isped.u-bordeaux2.fr/ISPED-EAD%20-20Guide%20utilisation%20de%20la%20plateforme.pdf>

*L'enseignement à distance en statistique et l'usage de logiciels au travers de deux exemples*

observés lorsqu'ils formulent leurs questions via un forum, mais il n'en est rien, car en réalité il y a une difficulté complémentaire qui consiste dans le fait de réussir à formuler sa question. La qualité des échanges au sein d'un forum de module ou d'UE dépend énormément de la qualité d'échanges qu'instaure un enseignant et du rythme des échanges, c'est-à-dire les délais de réponse.

Un forum a été ouvert spécialement pour aider les apprenants lors du téléchargement des différents logiciels, qui est une phase importante dans la mise en œuvre de la formation. L'accès au logiciel SAS pour les étudiants inscrits au master en première ou seconde année se déroule soit par un téléchargement directement à partir d'une adresse transmise par SAS ou exceptionnellement par envoi de 5 DVD si le téléchargement est impossible pour raison de très mauvaises connexions (pour certains étudiants il fallait plusieurs jours pour télécharger le logiciel). Une personne est donc en charge, via un forum, de veiller à la bonne installation du logiciel pour chaque étudiant, ce qui est source de nombreux échanges malheureusement pas toujours fructueux. Les téléchargements des logiciels EpiInfo et R se déroulent assez facilement pour la plupart des étudiants. Pour chaque logiciel un enseignant a été missionné comme responsable et a en charge de répondre à l'ensemble des questions directement liées à l'usage du logiciel.

Depuis quatre ans et suite à la demande d'étudiants, il a été décidé de proposer pour les DU des exercices supplémentaires avec des corrections individualisées ; cette mission est à la charge d'un attaché d'enseignement. La mise en place de ce nouvel outil a été fort appréciée par les étudiants et nous pouvons constater que les étudiants effectuant ces exercices ont de meilleurs résultats.

### **3.4.2 Statistique - Econométrie Toulouse**

L'ensemble de la formation est encadré par un tuteur accompagnant, un responsable administratif et un responsable pédagogique. Chaque cours est encadré par un tuteur spécialiste de la matière et qui peut être un enseignant-chercheur ou un professionnel. A l'heure actuelle, les enseignants des dix-sept cours existants se répartissent en 4 enseignants-chercheurs d'UT1C, 7 hors UT1C, 3 personnels INRA ou CNRS et 4 professionnels. L'équipe pédagogique et administrative se réunit une fois par an pour faire un bilan de l'année précédente et expliquer les nouveautés pour l'année à venir ainsi que pour présenter les nouvelles promotions.

Chaque cours dispose d'un espace Moodle que l'enseignant, tuteur du cours, gère à sa façon, mais il existe des espaces généraux de cours permettant aux étudiants de poser des questions d'ordre général et qui dans le passé ont permis de gérer le téléchargement et l'installation du logiciel SAS. Récemment, un espace de cours a été créé pour R et un autre pour SAS et tous les détails pour pouvoir réussir l'installation des ces logiciels se trouvent sur ces deux espaces de cours qui ont lieu au début de nos formations. Ainsi, le cours sur R contient plusieurs vidéos qui permettent aux étudiants de se familiariser avec l'organisation d'un cours mais aussi d'installer facilement R ainsi que l'éditeur de texte Tinn-R associé et de commencer la prise en main de R. Concernant SAS, nous avons aussi des détails sur le téléchargement et l'installation et lorsque des difficultés techniques surviennent auxquelles le tuteur de SAS ne sait pas répondre, nous nous adressons directement au support de SAS-Paris qui est heureusement très efficace. Depuis la rentrée 2011, le logiciel SAS n'est plus obligatoire, mais plus de la moitié des étudiants ont quand même suivi le cours sur SAS et l'ont téléchargé.

Nous constatons qu'il n'est pas facile de maintenir une activité régulière et féconde sur les forums malgré les exercices et projets à rendre régulièrement et les relances organisées par le tuteur accompagnant en charge de l'ensemble de la formation. Selon les promotions et pour des efforts complètement identiques de la part des enseignants, les forums peuvent être très actifs ou au contraire presque morts. Le problème essentiel vient de ce que certains apprenants sont trop occupés par ailleurs. La mise en place de l'année préparatoire a pour objectif non seulement de combler les lacunes de certains étudiants, notamment en mathématiques, mais aussi de pouvoir évaluer avant l'entrée en master 2 la capacité d'investissement dans les cours à distance des étudiants ; elle nous aide à sélectionner de bons étudiants dont on sait qu'ils auront la possibilité de dégager suffisamment de temps pour arriver au bout de la formation.

### 3.4.3 Point commun

Les deux offres de formation étant soutenues par l'AUF par le biais de bourses offertes à des candidats, elles se doivent de mettre à disposition en libre accès 25% du contenu pédagogique des formations soutenues par l'AUF. Le choix fait par Toulouse porte sur certains cours de Philippe Besse<sup>6</sup> et de Anne Ruiz-Gazen<sup>7</sup>. Tandis que l'ISPED étant soutenu par l'AUF pour un ensemble de formations dans le champ de la santé publique a préféré développer un site présentant un ensemble d'enseignements appelé « essentiels en santé publique »<sup>8</sup>.

Pour les deux formations il est très difficile actuellement d'assurer l'utilisation de SAS pour l'ensemble des étudiants car la difficulté majeure vient d'un téléchargement très long voire impossible pour les étudiants vivant dans des pays dont les connexions sont difficiles. Un étudiant indiquait sur le forum de l'ISPED : « J'ai mis 6 jours en continu 24/24 pour le téléchargement sur ce qui est la meilleure connexion à Lomé. »

## 3.5 Phase d'évaluation

### 3.5.1 Bordeaux

La mise en place de ce type de formation est compliquée et se doit d'être évaluée et réajustée régulièrement. Il est aussi important de mettre en place les outils les plus adéquats pour offrir la meilleure formation. Plusieurs étapes ont évolué depuis la première offre de formation via Internet, et grâce tout particulièrement aux retours que nous avons des étudiants par les enquêtes de satisfaction menées tous les ans à l'issue de la formation. La première modification a été de mettre en place des exercices supplémentaires avec des corrections individualisées, mais aussi de développer un chat ainsi qu'un partage de documents au niveau de la plate-forme d'enseignement pour aider les étudiants à travailler entre eux.

A la création de l'enseignement via Internet de l'ISPED, il n'avait pas été question de proposer R comme logiciel car il n'était pas si présent ni si utilisé par les professionnels qu'aujourd'hui. Mais depuis cinq ans, il est assez évident que sa place dans l'offre des logiciels pour des formations à distance est une évidence.

---

<sup>6</sup> <http://www.math.univ-toulouse.fr/~besse/enseignement.html>

<sup>7</sup> <http://www-gremaq.univ-tlse1.fr/stat/Annweb/sampling.htm>

<sup>8</sup> [http://campus.isped.u-bordeaux2.fr/PLEIADE/PV3/ASPX/PV2\\_VPS.aspx?16860](http://campus.isped.u-bordeaux2.fr/PLEIADE/PV3/ASPX/PV2_VPS.aspx?16860)

*L'enseignement à distance en statistique et l'usage de logiciels au travers de deux exemples*

Les différentes remarques des apprenants faites directement par messages dans le forum ou lors de l'enquête de satisfaction de fin d'année montrent qu'il leur est presque impossible d'apprendre à utiliser simultanément les logiciels R et SAS, et pour cette raison ils font un choix selon les différents critères. Un des éléments importants qui intervient dans ce choix est la possibilité d'avoir à proximité une personne utilisant déjà l'un ou l'autre des logiciels.

De plus, les enquêtes de satisfaction ont mis en évidence le fait que les étudiants construisaient des stratégies pour lutter contre l'isolement dans leur apprentissage : certains d'entre eux utilisent largement les forums liés directement à la formation, d'autres préfèrent intervenir dans d'autres forums ou ils préfèrent échanger directement entre eux par chat ou appel vidéo (skype par exemple). Ce type d'échange est particulièrement utilisé pour leurs questions concernant l'utilisation de logiciels car ils sont souvent en recherche d'obtenir une réponse rapidement et mettent donc en œuvre plusieurs démarches de recherche. Il faut souligner d'ailleurs le fait que les forums ayant pour thème les logiciels R et SAS sont très actifs.

### **3.5.2 Toulouse**

Après une ou deux années de tâtonnements pédagogiques, un mode de fonctionnement efficace a été trouvé et est amélioré d'année en année. Le forum est le lieu où chaque cours peut s'apprendre, le savoir se partager, les explications s'affiner. Les projets collectifs prennent parfois forme sur le forum, beaucoup de projets individuels s'y améliorent et les corrections d'exercices y sont discutées. Le succès du forum dépend de sa dimension collective. A une certaine période de l'année, une matière est enseignée, son forum est ouvert. C'est à ce moment-là qu'il faut étudier la matière, pour bénéficier de l'effet d'entraînement du groupe et aussi parce que les intervenants ne sont pas disponibles aux moments que choisirait chaque apprenant.

Depuis 5 ou 6 ans, nous avons mis en place un questionnaire d'évaluation pour chaque cours qui est à compléter par les étudiants en fin de trimestre. Le taux de réponse à ces questionnaires est d'environ 50% et la plupart des étudiants qui répondent ont suivi le cours de façon assidue. Les résultats de ces questionnaires qui sont positifs, voire très positifs, sont donc à considérer avec précaution. Un questionnaire général est aussi posé à la fin de la formation aux étudiants de master 2 et les apprenants qui s'expriment sont là aussi très satisfaits même s'ils ont en général des remarques sur des points à améliorer que nous cherchons à prendre en compte du mieux possible.

Enfin, l'analyse des candidats et des besoins a mis en évidence qu'il était important de proposer la création d'un DU et d'une année préparatoire au master 2, car les candidats au master 2 avaient parfois un manque de pré-requis et il n'existe pas de formations proposées via Internet leur permettant de combler ces lacunes.

### **3.5.3 Point commun**

Le suivi et l'évaluation du processus de formation via Internet est obligatoire car il y a tout d'abord une évolution rapide de la technique qui nécessite une mise à jour régulière des outils d'enseignement et des accès à l'information, comme par exemple le développement actuel des tablettes numériques. Ces évolutions nécessitent une réactivité importante pour les enseignants qui se doivent de mettre à jour leur support d'enseignement, ce qui nécessite un

travail lourd et coûteux en temps. Ces évaluations sont donc réalisées annuellement auprès des formateurs et des formés.

L'évaluation est dans tous les cas une phase inhérente à la formation.

Dans le cadre de ces formations, il ne faut pas oublier que l'évaluation passe aussi par les résultats aux contrôles continus et épreuves terminales. Les modalités d'évaluation correspondent à la remise de projets tout au long de l'année pour lesquels l'usage de logiciels est souvent nécessaire, mais aussi d'épreuves terminales par écrit qui se déroulent dans différents pays grâce au soutien de l'AUF.

## 4 Conclusion

L'offre de formation via Internet est particulièrement intéressante car elle offre à un public ayant des contraintes logistiques importantes l'opportunité de pouvoir suivre une formation de qualité et reconnue. Cependant, il est évident que ce type de formation est difficile à mettre en œuvre d'autant plus lorsqu'il s'agit, comme c'est le cas pour les formations qui ont été présentées, de toucher un public confronté à de lourdes contraintes organisationnelles.

L'un des points importants que nous avons souhaité soulever dans cet article est le questionnement que nous avons rencontré dans les deux cas pour définir et choisir les logiciels que nous allions privilégier pour l'apprentissage de la statistique à distance. Il n'y a pas de choix unique mais plusieurs points doivent être analysés conjointement : le profil des apprenants et leurs pré-requis en statistique, l'accessibilité à des connexions de qualité, l'investissement des enseignants dans le suivi de l'apprentissage des étudiants...

Il est souligné dans le cadre de nos formations qu'il est très important de ne pas penser une formation à distance uniquement comme une diffusion de savoir par des transmissions de supports d'enseignement, mais qu'il faut accompagner l'étudiant dans son apprentissage. Cet accompagnement nécessite donc une communication au moyen de forums mais aussi parfois d'échange de courriels plus individualisés.

En conclusion, ce type de formation est difficile à mettre en œuvre et nécessite un suivi de proximité avec les apprenants, mais aussi de veille pour ce qui concerne les nouvelles technologies. Cela implique une forte implication de la part d'un ensemble d'acteurs : enseignants, ingénieurs pédagogiques, informaticiens et administratifs.

## Références

- [1] Aragon, Y., T. Laurent et A. Ruiz-Gazen (2007), Enseignement de la statistique à distance : l'expérience du M2 Statistique & Econométrie, MoodleMoot, Castres.
- [2] Härdle, W., S. Klinke, and U. Ziegenhagen (2006), E-Learning Statistics - A Selective Review, SFB 649, Humboldt-Universität zu Berlin, <http://sfb649.wiwi.hu-berlin.de>.
- [3] Molenda, M., J. A. Pershing, and C. M. Reigeluth (1996), Designing instructional systems, in R. L. Craig (Ed.), The ASTD training and development handbook 4<sup>th</sup> ed., New York, 266-293.