



OFFRE DE FORMATION ED211

ÉCOLE DOCTORALE SCIENCES EXACTES ET LEURS APPLICATIONS - ED 211

Avenue de l'université BP 1155 64013 PAU Cedex – France

Intitulé de la formation

Logiciel R (enseignement à distance)

- **Période : non pertinent (car enseignement à distance). Durée de la formation : 6 ou 12 semaines.**
- **Nb d'heures : 6 heures en présentiel**
- **Responsable de l'enseignement: Ch. Paroissin**

PRÉ-REQUIS :

Notions de statistique en liaison avec le contenu de la formation.

CONTENU DE L'ENSEIGNEMENT :

L'objectif de la formation vise à l'apprentissage du logiciel R pour l'analyse statistique. Ce logiciel est libre, gratuit et disponible sur tous les systèmes d'exploitation. Il est de plus en plus utilisé dans diverses domaines de la recherche, des sciences dures aux sciences humaines.

Le contenu de l'enseignement dépend des besoins du doctorants. Il doit choisir 12 modules dans une liste de modules disponibles (voir plus bas). Le choix des modules se fait en concertation entre le doctorant et le responsable de cette formation (voir plus loin à la rubrique « Organisation pédagogique »).

Attention, il ne s'agit pas d'un cours de statistique, mais d'un cours de mise en œuvre de techniques statistiques à l'aide du logiciel R. Les rappels sur les notions de statistique utiles seront donc minimalistes et parfois insuffisants. Cette formation vient donc en appui de cours de statistique que le doctorant aura suivi au préalable (le forum du cours en ligne pourra néanmoins être utilisé pour poser des questions à ce niveau-là).

Les modules sont regroupés en deux thématiques : « langage de programmation » et « analyse statistique ».

Liste des modules disponibles d'ici janvier 2017 au niveau « langage de programmation » :

- Création et manipulation d'objet
- Gestion des entrées/sorties
- Éléments de programmation
- Créer ses propres fonctions
- Fonctions mathématiques
- Réaliser des (jolis) graphiques

Liste des modules disponibles d'ici janvier 2017 au niveau « analyse statistique » :

- Statistiques descriptives univariées
- Statistiques descriptives bivariées
- Lois de probabilité : calculs, simulations et estimation paramétrique pour un échantillon
- Méthodes non-paramétriques
- Tests de comparaison d'échantillons gaussiens
- Tests non-paramétriques de comparaison
- Test d'adéquation
- Série temporelles : techniques de lissage
- Régression linéaire : techniques de base

- Analyse de la variance et de la covariance
- Analyse de données multivariés (ACP, AFC, ACM)
- Méthodes de classification
- Indépendance et corrélation
- Cartographie

Dans un avenir plus lointain, d'autres modules seront développés tels que :

- Combiner R avec C ou Fortran
- Création d'une interface Shiny
- Rapports automatiques avec OdfWeave ou Sweave
- Débogage de programme
- Création de package
- Séries temporelles : modélisation SARIMA
- Analyse statistique de durées de vie : modèles paramétriques
- Analyse statistique de durées de vie : méthodes et modèles non-paramétriques
- Régression linéaire : approfondissement

D'autres modules pourraient également être développés selon les besoins des apprenants (à condition que la demande soit exprimée suffisamment à l'avance).

COMPÉTENCES VISÉES :

Maîtrise de certaines fonctionnalités du logiciel R. Autonomie (partielle ou totale) pour mener à bien une analyse statistique dans le cadre de leur thèse.

ORGANISATION PÉDAGOGIQUE :

La formation commence par une première séance collective de 2 heures durant laquelle la formation est présentée au début. Les doctorants seront alors ensuite invités à présenter leurs problématiques et le type de données qu'ils devront analyser. Ces présentations serviront à définir les modules que chaque doctorant devra suivre, certains modules pouvant être communs.

La formation en elle-même est dispensée à distance. Il est prévu un rythme d'un module par semaine, ce qui aboutit à une durée totale de la formation de 12 semaines. Néanmoins la formation peut être dispensée de manière accélérée sur 6 semaines (2 modules par semaine).

Chaque module est constitué de notes de cours suivi par un test d'entraînement. Ce test ne rentre pas en compte dans la note, mais il est nécessaire d'avoir un score d'au moins 75 % pour accéder au module suivant. Des tutoriels vidéo pourront compléter les notes de cours (en cours de réalisation).

Il est prévu deux autres séances de 2 heures, l'une à mi-parcours et l'autre à la fin de la formation. Ces séances permettront aux doctorants de s'exprimer (bien qu'un forum soit disponible sur la formation en ligne) et de faire de mises au point. La dernière séance sera aussi utile pour dresser un bilan de la formation.

La mise en place de cette formation en ligne bénéficie du soutien (logistique, technique, pédagogique, etc.) du pôle ARTICE (DN, UPPA) via l'appel FOAD/UPPA et Pyren.

A noter que certains modules sont disponibles en espagnol.

APPROFONDISSEMENTS POSSIBLES :

Dans le cadre d'une autre formation au logiciel R, les doctorants pourront suivre une formation complémentaire, sur le même principe, en choisissant 12 modules différents de la première formation.