

# Stat. et Science des données pour la biologie I



Niveau  
d'étude  
BAC +5 /  
master



ECTS  
5 crédits



Composante  
Faculté des  
sciences

## En bref

- › Langue(s) d'enseignement: Français
- › Ouvert aux étudiants en échange: Oui

## Présentation

### Description

Cet enseignement, divisé en trois parties respectivement intitulée Introduction à l'Analyse de Survie, Introduction aux plans d'expériences et Modèles de reconstruction de généalogie, a pour objectif de présenter quelques thèmes importants de Statistique et Modélisation pour la biologie. Ci-dessous une description plus précise.

#### Introduction à l'Analyse de Survie

Le problème des données manquantes est fréquemment rencontré dans un grand nombre de domaines d'applications statistiques. Dans le cadre de la biologie, il n'est pas rare que l'acquisition de données de suivi soit en partie incomplète : pour un essai thérapeutique, le temps de rémission puis de guérison d'un patient peut être trop long et donc jamais observé, on parle alors de données censurées.

Dans ce cours, nous introduirons la méthodologie (fonction de survie, risque instantané, ...) et les outils statistiques (estimateur de Kaplan-Meier, Nelson-Aalen, ...) qui permettent d'appréhender ce genre de problèmes. Une implémentation sur R sera aussi proposée.

#### Introduction aux plans d'expérience

Les plans d'expérience forment un outil essentiel en statistique expérimentale, ils permettent d'esquisser les contours d'une méthodologie et d'organiser au mieux la suite des mesures à effectuer afin de répondre à une problématique statistique donnée.

Nous décrirons de manière succincte quelques plans d'expérience classiquement utilisés dans les sciences du vivant.

#### Modèles de reconstruction de généalogie

Ce module est dédié à l'étude d'un cas pratique de reconstruction généalogique dans une population végétale. Après quelques concepts génétiques généraux liés aux niveaux de ploïdie, à la production des gamètes et aux croisements, nous proposerons un modèle permettant de reconstruire, à partir de marqueurs génétiques, un ensemble d'arbres généalogiques probabilisés dans une population végétale soumise à des lois de reproduction

spécifiques. Nous étudierons des reconstructions dans des populations diploïdes ou tétraploïdes, et un travail pratique sera demandé dans le cas diploïde.

Une partie de ce travail sera dédiée à la programmation effective de la reconstruction généalogique en R ou en Python. La visualisation des résultats sera réalisée à l'aide du package `igraph` de R ou de son équivalent en Python.

## Objectifs

Logit-Probit-Tobit-Modèles de Heckman. Choix discrets/ Les modèles de choix binaire (probit)/modèle Logit/modèle Logit multinomial/modèle Logit conditionnel/ modèles à choix ordonnés/modèles de comptage/modèles à variables limitées (tronquées)/modèles d'autosélection/ Arbres de décision et Nested Logit. Visualisation de graphes. Arbres probabilisés. Principe des plans d'expérience.

## Heures d'enseignement

CM	Cours magistral	16h
TP	Travaux pratique	20h

## Pré-requis obligatoires

*Notions et contenus :*

- Cours de introduction à la génomique, Statistique et Modèles de régression du MI.
- Compétences en R et/ou Python.

*Compétences et capacités :*

Maîtriser les concepts généraux des probabilités, des statistiques et des modèles de régression vus en MI, ainsi que l'introduction à la génomique.

## Infos pratiques

---

Lieu(x)

> Angers