

# Licence Mathématiques et informatique appliquées aux sciences humaines et sociales

## Statistique avec le logiciel R

### Informations

Composante : Faculté des Sciences

### Responsable

Annie BROGLIO (enseignant-)

### Langue(s) d'enseignement

Français

### Contenu

Cours prévu en deux parties: apprentissage des rudiments du logiciel R et travail sur un projet personnel

1. Le logiciel R: Installation et configuration de R, RStudio, RMarkdown, les objets R : vecteurs, tableaux, matrices, dataframe, collecte et nettoyage des données, importation et exportation de données.

Bases de la programmation en R, créer ses propres fonctions, étude de certaines bibliothèques, par exemple dédiées au graphisme, ggplot2, à la science des données tidyverse etc...

Visualisation de données: statistiques descriptives uni et multivariées, cartographie, simulations, lois classiques, tirages aléatoires. Les fonctions R pour la regression linéaire simple, les tests statistiques sur les proportions, moyennes, variances.

2. Projet de sciences de données, sur des données réelles choisies par l'étudiant: de la collecte des données, mise en place d'un questionnaire, à l'analyse et à la conclusion. Utilisation de techniques vues dans la première partie du cours, éventuellement complétées par un travail spécifique si nécessaire. Une ré-utilisation des techniques vues en atelier Maths-Shs en parallèle est aussi possible ( tests de normalité, anova, contrastes...)

**Remarque:** Plusieurs TP « types » sont fournis pour servir de modèle dans la première partie du cours, pour des analyses de données ultérieures, on aborde les notions de fiabilité et nettoyage des données, les problèmes de données manquantes, ...

### Modalités d'organisation

18h de cours 32 h de tp

### Bibliographie, lectures recommandées

Le logiciel R: Maîtriser le langage, effectuer des analyses (bio)statistiques

[Pierre Lafaye de Micheaux](#) , [Rémy Drouilhet](#) , [Benoît Liqueur](#) Springer

Comprendre et réaliser les tests statistiques à l'aide de R Gael Millot De Boeck

R for Data Science de Hadley Wickham et Garrett Golemund

Introduction to Statistical Learning with R de Gareth James, Daniela Witten, Trevor Hastie et Robert Tibshirani

Tutoriels : RStudio Learning Center, DataCamp,...

### Pré-requis obligatoires

Estimation et tests

probabilités et statistiques 1 et 2

### VOLUME HORAIRE

- Volume total: 50 heures

- Cours magistraux: 18 heures
- Travaux pratiques: 32 heures

### Codes Apogée

- SMH5U13C [ELP]

### Pour plus d'informations

[Aller sur le site de l'offre de formation...](#)



Dernière modification le 07/10/2024