

## Formations Intensives de l'École HED

### Formations R (2022)

**Formateurs :** Bénédicte Garnier, Elisabeth Morand, Marc Thevenin

Entre avril et juin 2022, 3 sessions de formation R, du niveau Débutant au niveau Avancé seront organisées dans le cadre des *Formations intensives de l'École des Hautes Études en Démographie* (EUR HED) :

- Manipuler ses données et programmer avec R • niveau débutant (11 & 12 avril 2022) ;
- Analyses statistiques avec R • niveau avancé (Programmée en mai 2022, la date sera communiquée ultérieurement)
- Approches longitudinales avec R • niveau avancé (Programmée en juin 2022, la date sera communiquée ultérieurement)

Les formations peuvent être suivies de manière indépendante à condition d'avoir les connaissances suffisantes (cf prérequis).

**Il ne s'agit pas de cours de statistiques :** les différents modules présupposent donc que vous avez déjà une connaissance des différentes techniques présentées.

#### Lieu de formation :

Campus Condorcet

Institut National d'Études Démographiques

9 cours des Humanités,

Salle 3 .030

93322 AUBERVILLIERS CEDEX,

**Information :** Pour ces 3 ateliers, vous pouvez apporter votre ordinateur portable.

**NB :** *Les formations Intensives de l'École HED* peuvent délivrer des crédits ECTS. Une attestation de présence pourra vous être délivrée à l'issue de la formation ou sur demande.

Nous invitons chaque doctorant à se renseigner auprès de la direction de son école doctorale.

Merci de retrouver le descriptif de chacune des formations ci-dessous.

Information et inscriptions: [formations-intensiveshed@ined.fr](mailto:formations-intensiveshed@ined.fr)

## Manipuler ses données et programmer avec R • niveau débutant

11 au 12 avril 2022

### Objectif :

Présenter comment manipuler des données et diverses opérations courantes (comme construire des données ou produire des graphiques) avec R.

### Thématiques abordées :

- Prise en main de R et RStudio
- Importer un fichier de données
- Manipulation des données et recoder
- Discrétiser une variable continue
- Récupérer un sous ensemble de données
- Identifier et compter des données manquantes, identifier des individus atypiques
- Produire des statistiques descriptives uni et bivariées (incluant intervalles de confiance et tests)
- des distributions statistiques uni variées ; des caractéristiques de position (moyenne, médiane, quantiles...) et de dispersion (variance, écart-type...)
- Fusionner des données
- Agréger des données
- Mettre en forme des données répétées (format *long* et *wide*)
- Gestion des dates : le *package lubridate*
- Introduction à *ggplot2* (graphiques)
- Produire des rapports automatisés avec *knitr* et *rmarkdown*

### Prérequis :

- Pas de connaissance de R préalable
- Connaître les statistiques uni- et bivariées et la régression logistique
- Une connaissance d'un autre logiciel de statistique (tel que SPSS, SAS ou Stata) est préférable

### Acquis de la formation :

Combiner différentes fonctions dans un même programme pour manipuler ses données  
Réaliser les graphiques adaptés les plus courants

## Analyses statistiques avec R • niveau avancé

Programmée en mai 2022, la date sera communiquée ultérieurement

### Objectif :

L'objectif de ce second atelier est de perfectionner son usage de R et RStudio en abordant des techniques statistiques plus avancées (analyse des correspondances, analyse multivariée...) qui nécessitent le plus souvent une manipulation plus complexe des données en amont.

### Thématiques abordées :

- Rappel sur les distributions statistiques uni variées et les caractéristiques de position et de dispersion
- Tableau de contingence
- Édition d'indicateurs de liaison entre variables (statistique du khi-deux, V de Cramer, coefficient de corrélation linéaire...)
- ODDS, ODDS ratio, Risque Ratio
- Plan d'échantillonnage et pondération
- Analyse des données et classifications
- Régression logistique
- Sélection *pas à pas* d'un modèle
- Modèles multivariés (approches *stepwise*, effets d'interaction, multicollinéarité...)
- Présentation d'autres modèles : modèle de poisson, régression multinomiale, modèles mixtes

### Prérequis :

- Connaissance du logiciel R et de sa syntaxe (cf. contenu de la formation Manipuler ses données et programmer avec R • niveau débutant)

### Acquis de la formation :

- Produire des statistiques avec le logiciel R
- Mettre en œuvre sous R des analyses statistiques avancées

## Approches longitudinales avec R et RStudio • niveau avancé

Programmée en juin 2022, la date sera communiquée ultérieurement

### Objectif :

Introduction aux approches longitudinales dont analyse des durées, analyse de séquences et *optimal matching* dans le champ des sciences sociales et à son implémentation sous R.

### Thématiques abordées :

- Introduction à l'analyse des durées
  - Données et théorie (données censurées et grandeurs théoriques)
  - Analyse non paramétrique (Estimation des fonctions de survie et tests de différence)
  - Modèles à risques proportionnels (modèles semi-paramétrique et à temps discret)
- Introduction à l'analyse de séquences

### Prérequis :

- Connaissance du logiciel R et de sa syntaxe (cf. contenu de la formation Manipuler ses données et programmer avec R • niveau débutant)
- Manipulation avancée de données et gestion des dates (cf. contenu de la formation Analyses statistiques avec R • niveau avancé)
- Savoir réaliser un modèle multivarié sur R et une classification hiérarchique (cf. contenu de la formation Analyses statistiques avec R • niveau avancé)
- 

### Acquis de la formation :

- Mettre en œuvre sous R des analyses longitudinales