

Durée : 3 jours soit 21 heures

Référence : IF-LANG-R

Public visé :

- Utilisateurs finaux, statisticiens, analystes type **Data Miner**, acteurs impliqués dans l'analyse/fouille des données.

Pré-requis :

- Connaissances de base en statistiques (régressions, analyse en composantes principales, classification) et des concepts de programmation.

Objectifs pédagogiques :

- Installer l'environnement d'analyse **R**
- Utiliser la console de **R**, et les environnements de travail **R Studio**
- Créer et regrouper plusieurs types d'objets de **R**
- Créer des programmes d'analyse avec **R**
- Élaborer des graphiques avec **R**
- Utiliser les packages de **R** pour mettre en œuvre des modélisations statistiques (**régression, ACP...**)

Compétences acquises à l'issue de la formation :

- Installer l'environnement d'analyse **R**
- Utiliser la console de **R**, et les environnements de travail **R Studio**
- Créer et regrouper plusieurs types d'objets de **R**
- Créer des programmes d'analyse avec **R**
- Élaborer des graphiques avec **R**
- Utiliser les packages de **R** pour mettre en œuvre des modélisations statistiques (**régression, ACP...**)

Modalités pédagogiques :

Session dispensée en présentiel ou téléprésentiel, selon la modalité inter-entreprises ou intra-entreprises sur mesure.

La formation est animée par un(e) formateur(trice) durant toute la durée de la session et présentant une suite de modules théoriques clôturés par des ateliers pratiques validant l'acquisition des connaissances. Les ateliers peuvent être accompagnés de Quizz.

L'animateur(trice) présente la partie théorique à l'aide de support de présentation, d'animation réalisée sur un environnement de démonstration.

En présentiel comme en téléprésentiel, l'animateur(trice) accompagne les participants durant la réalisation des ateliers.

Moyens et supports pédagogiques :

**Cadre présentiel**

Salles de formation équipées et accessibles aux personnes à mobilité réduite.

- Un poste de travail par participant
- Un support de cours numérique ou papier (au choix)
- Un bloc-notes + stylo
- Vidéoprojection sur tableau blanc
- Connexion Internet
- Accès extranet pour partage de documents et émargement électronique

**Cadre téléprésentiel**

Session dispensée via notre solution iClassroom s'appuyant sur Microsoft Teams.

- Un compte Office 365 par participant
- Un poste virtuel par participant
- Un support numérique (PDF ou Web)
- Accès extranet pour partage de documents et émargement électronique

Informations sur l'accessibilité :



## Description / Contenu

### Formation Environnement R, traitement de données et analyse statistique.

**R** est un **langage de programmation** dont le but est de pouvoir traiter et organiser des jeux de données en y appliquant des tests **statistiques** plus ou moins complexes. **R** permet de se représenter ces données traitées graphiquement à l'aide d'une grande variété de graphique disponible. Développé initialement au début des années 1990 par Ross Ihaka et Robert Gentleman, ce langage basé sur un autre **langage statistique appelé S** (oui, les informaticiens ont beaucoup d'imagination), est désormais maintenu et mis à jour par une équipe de développeurs au sein du **R Project**. Cette structure garantit des mises à jour fréquentes et une communauté importante d'utilisateurs apporte son aide pour notamment développer de nouvelles fonctionnalités (des librairies, nous en parlerons plus tard) au projet.

Le **langage R** est en effet très implanté dans la communauté scientifique et universitaire et ses nombreux utilisateurs organisent même des conférences pour discuter des nouveautés disponibles sous **R**.

#### Module 1 : Introduction

- Présentation du logiciel R et de ses fonctionnalités.
- Avantages et inconvénients.
- Accès au site de téléchargement de l'outil et installation. Exercice Installation de l'environnement d'analyse.

#### Module 2 : Premiers pas

- Environnement de base (console, script).
- Utilisation de la console.
- Création et sauvegarde d'un script.
- Le répertoire sous R Installation.
- Aide et commentaires.
- Les autres éditeurs Tinn-R et R Studio.
- Exercice *Création de scripts*.

#### Module 3 : Objets et notions de programmation R

- Les objets de type vecteurs, matrix, array, factor, data. frame, list.
- Manipulation des objets, classe d'objet, fonctions spécifiques, jointure.
- Sauvegarde, suppression de mémoire.
- Notions de boucle (for et while), condition (if), switch.
- Exercice *Manipulation des types d'objets*.

#### Module 4 : Création et utilisation de fonctions

- Structure d'une fonction.
- Fonctions de type mathématique.
- Fonctions de type chaîne de caractères.
- Fonctions liées au temps/date.
- Opérations ensemblistes.
- Les tables de contingences.
- Exercice *Création et utilisation de fonctions*.

#### Module 5 : Génération, gestion et visualisation des données

- Les données : séquences régulières et aléatoires.
- Données exemples de R.
- Importation et exportation de données.
- Modifier les données d'un objet.
- Exemples de graphiques construits avec R.
- Création des graphiques de base.
- Les options graphiques, partager une fenêtre graphique, sauvegarder un graphique.
- Exercice *Exercice d'application sur les données*.

#### Module 6 : Analyses statistiques

- Présentation de la notion de package (librairie).
- Télécharger/charger des packages.
- Quelques packages utiles.
- Cas de la régression non linéaire multiple.
- Cas de l'analyse en composantes principales ACP.
- Cas de la classification CAH.
- Exercice *Exercice d'application sur les packages*.