

Langage R - Traitement de données et analyses statistiques

2 j (14 heures)

Ref : LAGR

Public

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

Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue
Exposés, cas pratiques, synthèse, assistance post-formation pendant trois mois
Un poste par stagiaire, vidéoprojecteur, support de cours fourni à chaque stagiaire

Modalités de suivi et d'évaluation

Feuille de présence émargée par demi-journée par les stagiaires et le formateur
Exercices de mise en pratique ou quiz de connaissances tout au long de la formation permettant de mesurer la progression des stagiaires
Questionnaire d'évaluation de la satisfaction en fin de stage
Auto-évaluation des acquis de la formation par les stagiaires
Attestation de fin de formation

Objectifs

- Installer l'environnement d'analyse R
- Utiliser la console de R, et les environnements de travail Tinn-R et R Studio
- Créer et regrouper plusieurs types d'objets de R
- Créer des programmes d'analyse avec R
- Elaborer des graphiques avec R
- Utiliser les packages de R pour mettre en oeuvre des modélisations statistiques

Programme détaillé

ENVIRONNEMENT R : INTRODUCTION

Présentation du logiciel R et de ses fonctionnalités

- Avantages et inconvénients
- Accès au site de téléchargement de l'outil et installation
- 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

OBJETS ET NOTIONS DE PROGRAMMATION R

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

CREATION 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

GENERATION, GESTION ET VISUALISATION DES DONNEES

- 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

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