



Programmation statistique avec R

Référence et durée

R_STAT

2 à 3 jours

Objectif

Ce cours se propose de faire découvrir les principales fonctionnalités statistiques et datamining offertes par le logiciel R et ses bibliothèques dédiées : de l'analyse exploratoire à la modélisation explicative et prédictive

Public et Prérequis

Chargés d'étude, statisticiens, biostatisticiens, data analysts, data scientists

Avoir suivi le cours R_DATA ou avoir un niveau de connaissances équivalent. Avoir des connaissances théoriques en statistiques et modélisation.

Programme

- Principaux packages par fonctionnalités (analyse de données, régressions, analyse prédictive, tests, visualisation graphique ...)
- Visualisation et Exploration des données sous R
 - L'Analyse graphique : diagrammes en barres, nuages de points, boîte à moustaches...
 - Tris croisés, agrégations par groupes, statistiques descriptives
 - La recherche de liaisons entre variables : Corrélation, Test du khi2, test de comparaison de moyennes (Student, Wilcoxon, Anova...), comparaison de variance (Fisher, Bartlett), Test de normalité ...
- Analyse de Données multidimensionnelles
 - Analyse Factorielle, Analyse en Composantes Principales
 - Partitionnement par kmeans, classification ascendante hiérarchique, classification mixte, caractérisation
- Modélisation de données explicative et prédictive sous R
 - L'analyse de variance et de covariance
 - Les modèles de régression linéaire multiple : spécification, détection des points atypiques, colinéarité, sélection de variables
 - La régression logistique : spécification, sélection de variables, qualité prédictive
- Arbres de décision
 - Construction de l'arbre de classification
 - Construction de l'arbre de régression