



Logiciel R niveau II

Statistiques avancées

du 8 au 10 novembre 2016 - Marseille

Public	<p>Chercheurs, ingénieurs et techniciens souhaitant se perfectionner dans l'utilisation du logiciel R dans un contexte de mise en œuvre d'analyses statistiques des données. Ces utilisateurs de R viennent chercher du savoir-faire logiciel sur la mise en œuvre des analyses statistiques et le travail sur les tableaux de données. Il ne s'agit donc pas d'une formation sur le fond statistique mais bien d'une formation orientée logiciel. Cette formation n'est pas adaptée aux apprenants à la recherche d'un savoir-faire sur la programmation avancée sous R (boucles, tests conditionnels, création de fonctions d'automatisation performantes...)</p> <p>Prérequis</p> <ul style="list-style-type: none"> > Connaissances de base du logiciel R indispensables : création et manipulation des objets sous R, importation des données, utilisation simple des commandes de R en mode script ou en mode console. > Connaissances indispensables sur les différentes analyses statistiques mises en œuvre lors de la formation
Objectifs	<ul style="list-style-type: none"> • Lire et écrire des scripts « avancés » d'analyse des données • Mettre en œuvre et extraire les résultats aussi bien d'une analyse de la variance à 1 facteur que des tests post-Hoc de comparaisons multiples • Mettre en œuvre et extraire les résultats d'une analyse de la variance multi-facteurs • Mettre en œuvre et extraire les résultats d'une régression linéaire simple et multiple • Mettre en œuvre et extraire les résultats d'une analyse multidimensionnelle des données (ACP)
Intervenante	Sandrine Longis - Arkesys
Lieu	CNRS-GLM, 31 chemin Joseph Aiguier, 13009 Marseille
Contact	Marie-Laure Olive, responsable formation Inserm - demat-form.dr-marseille@inserm.fr
Inscriptions	<ul style="list-style-type: none"> • Personnels travaillant dans une structure mixte Inserm : inscription via www.sirene.inserm.fr Région : Paca Domaine : TS4 - Statistiques / épidémiologie • Personnels CNRS : formulaire d'inscription papier à transmettre à formation@dr12.cnrs.fr • Autres personnels : formulaire papier à transmettre à demat-form.dr-marseille@inserm.fr

Date limite d'inscription : 30 septembre 2016

Logiciel R niveau II - Statistiques avancées - du 8 au 10 novembre 2016 - Marseille

PROGRAMME DETAILLE

Revoir les fondamentaux sur le logiciel R

Fonctionnement général du logiciel et documentation en ligne
Conseils sur l'organisation d'un projet R
Affichage, sauvegarde, suppression et chargement des objets en mémoire
Importation des données contenues dans un classeur Excel, un fichier texte ou CSV
Ouvrir une boîte de dialogue de gestion de fichiers
Notions d'objet, de package et de répertoire courant
Installation et utilisation d'un package
Utilisation d'interfaces type Rcmdr
Trucs et astuces

Créer et manipuler des objets sous R

Attributs intrinsèques des objets
Notion de vecteur, facteur, matrice, liste
Création et indexation des vecteurs (variables quantitatives)
Création et manipulation des facteurs (variables qualitatives)
Convertir l'attribut d'un objet
Extraction des éléments d'une sortie R

Structurer et extraire de l'information d'un tableau de données

Création et manipulation efficace des variables qualitatives et quantitatives
Ajout d'observation ou de variables à un tableau de données
Concaténer des colonnes
Empilement / dépilement d'un tableau de données
Gestion des données manquantes
Création de tableaux de synthèse (tableaux croisés)
Création de tableaux de contingences

Revoir la mise en œuvre des analyses statistiques de base sous R

Analyse descriptive des données
Intervalles de confiance
Tests d'hypothèse paramétriques et non paramétriques
Tests paramétriques de comparaison de deux échantillons (Student, Welch)
Tests non paramétriques de comparaison de deux échantillons (Wilcoxon, Mann-Whitney)
Comparaison des variances (Fisher)
Test de normalité (Shapiro-Wilk) et QQ-plot
Comparaison de proportions (χ^2)
Corrélation linéaire et non paramétrique (Bravais-Pearson, Spearman)
Rappels et mise en œuvre de graphiques (en écrivant les commandes et avec des interfaces : Rcmdr et GrapheR)

Mettre en œuvre une ANOVA sous R

Les commandes `aov()` et `lm()`
Mise en œuvre d'une ANOVA à un facteur
Tableau d'analyse de la variance et estimation des coefficients du modèle
Vérification des conditions d'utilisation du modèle : normalité et homogénéité des dispersions (test de Bartlett, test de Levene)
Représentation graphique des résultats de l'analyse
Tests post-Hoc (comparaisons multiples)
Les commandes de base (TukeyHSD, `pairwise.t.test`)
Package spécifique (multcomp)
Tests non paramétriques : Kruskal-Wallis, Friedman
Mise en œuvre d'une ANOVA multi-facteurs
Éléments de syntaxe : *, |, /, Error...
Effets simples, effets croisés
Graphique des interactions
Somme des carrés de type I, de type II et de type III

Mettre en œuvre une régression linéaire sous R

Écriture d'un modèle : les fonctions, les opérateurs
Régression linéaire simple
Régression linéaire multiple
Tableau d'analyse de la variance et estimation des coefficients du modèle
Vérification des conditions d'utilisation d'une régression linéaire
Normalité des résidus
Homogénéité des résidus
Prédiction et intervalles de confiance d'une prédiction
Diagnostic des problèmes de multicolinéarité
Régression linéaire pas à pas

Réaliser une analyse factorielle multivariée sous R

Présentation des principaux packages
ACP (Analyse en Composantes Principales)
Réaliser une ACP avec le plug-in FactoMineR de Rcmdr
Identification des sorties :

- Valeurs propres et diagramme des valeurs propres
- Contributions et \cos^2 des variables
- Contributions et \cos^2 des individus
- Cercle factoriel
- Graphique des individus

Mettre en œuvre d'autres analyses statistiques - sur demande et si le temps le permet. Par exemple : analyse de covariance, analyse de variance pour mesures longitudinales (répétées), modèle mixte, analyse factorielle des correspondances (AFC), analyse factorielle des correspondances multiples (ACM), classification ascendante hiérarchique, modèle linéaire généralisés, etc.

Outils : il sera possible d'utiliser en plus du logiciel R et des bibliothèques de bases, l'interface graphique Rcmdr (R Commander), son plug-in FactoMineR, l'interface graphique GrapheR mais aussi le logiciel RStudio.

Méthode : après avoir revu les bases du logiciel, des méthodes statistiques seront mises en œuvre avec le logiciel de manière plus poussée qu'à l'occasion du module d'initiation et en suivant l'ordre logique du traitement statistique d'un tableau de données. Le choix de la méthode entre Rcmdr ou l'écriture des commandes ainsi que le choix d'un éditeur de script (R ou l'IDE RStudio par exemple) sera individuel et se fera selon les objectifs et les préférences de chacun. Les commandes seront analysées de sorte que les participants se familiarisent progressivement au code et qu'ils soient capable de les adapter lorsque cela sera nécessaire. L'objectif étant qu'ils soient capables de créer un script relatif à une étude de données et que ce dernier soit réutilisable par la suite, sans avoir à réutiliser Rcmdr ou saisir les commandes.

La pratique dirigée sera omniprésente et ponctuée d'exercices en autonomie. Les corrections tiendront compte des choix individuels quant à l'utilisation des interfaces.