

Journées d'analyse statistique des données sur R Edition 2022 délocalisée à Barcelonnette (Sud Alpes)

28 mars – 1^{er} avril 2022

Pôle Universitaire SEOLANE
Quartier du 11ème BCA - 04400 Barcelonnette

www.seolane.org



Journées R ECCOREV 2016 à Barcelonnette

Information : Joëlle Cavalieri, FR ECCOREV, mail : joelle.cavalieri@univ-amu.fr

La formation est ouverte en priorité aux doctorants et jeunes chercheurs des UMR affiliées à la fédération ECCOREV, aux Instituts d'établissement ITEM & OCEAN.

Lien site ECCOREV liste des laboratoires fédérés : https://www.eccorev.fr/IMG/pdf/liste_unites_eccorev_dec_2020-2.pdf

Contexte COVID-19 et considérations pratiques. La prise en charge comprend le transport à Barcelonnette depuis Marseille Saint Jérôme (stationnement voiture possible), l'hébergement en chambres collectives et la restauration sur place. Des navettes vers le centre-ville de Barcelonnette pourront être organisées. Une après-midi récréative impose de s'équiper pour la randonnée en montagne. Veillez à disposer d'un ordre de mission établi par votre employeur. Un engagement à participer à l'ensemble des sessions conditionne l'éligibilité au plan de formation doctorale. *Un protocole sanitaire sera établi pour la durée de l'événement.*

Ces journées se dérouleront sur ordinateur, avec le logiciel R, qui est un logiciel libre d'usage très répandu dans la communauté scientifique « Eccorev ». Le conférencier présentera les techniques en les illustrant avec des exemples reproductibles par les étudiants immédiatement sur leur ordinateur. Il y aura donc à la fois vidéo-projection et TP en parallèle.

L'inscription se fera pour l'ensemble des sessions. La salle peut accueillir **25 participants** en présentiel. Chaque participant amène son ordinateur portable personnel équipé du wifi. La configuration logicielle sera précisée avant le début des journées et nous anticiperons l'installation des packages pour chaque session. Le public visé (niveau débutant) sera en priorité : les thésards, post-docs et jeunes chercheurs des UMRs membres d'ECCOREV, des Instituts ITEM & OCEAN, en particulier ceux inscrits à l'ED251 selon la place disponible. L'objectif est de rendre l'utilisateur suffisamment autonome dans les domaines les plus « populaires » de la statistique, de manière qu'il soit capable de pratiquer directement les cas standards et d'être capable de personnaliser son approche.

L'équipe de formateurs comprend : Alberte Bondeau (CNRS-IMBE), Joël Guiot (CNRS Emérite – Cerege), Maxime Logez (INRAE – RiverLy), Laura March (IRD – LPED) & Franck Torre (AMU-IMBE).

J1 – Lundi 28 mars 2022

J1 – matin

Transport vers Barcelonnette

J1 – 13h30-16h30

Introduction à R

Par Franck TORRE, AMU-IMBE & Maxime LOGEZ, INRAE-RiverLy

Le but de cette session est de permettre à l'utilisateur novice de naviguer au sein l'environnement et d'utiliser les outils offerts par R pour l'analyse de données. Quelques applications graphiques et statistiques simples seront travaillées.

J2 – Mardi 29 mars 2022

J2 – 9h30-12h30

Analyse de données environnementales multivariées

Par Franck TORRE, AMU-IMBE

L'analyse de données permet de mettre en évidence l'information contenu dans un tableau de données multivariées. En fonction de la nature de ces variables, différentes méthodes ont été proposées et leur présentation est au programme de cette séance : analyse en composantes principales normées ou centrées, analyse factorielle des correspondances, analyse des correspondances multiples, analyses de données mixtes. Les méthodes de couplage de tableaux (coïnertie, RDA/CCA) permettent d'étudier le lien entre deux tableaux. Ces dernières permettent de décomposer la variance d'un tableau à expliquer selon différents compartiments de variables explicatifs. Des exemples provenant d'échantillonnage en écologie serviront d'illustration : tableaux biologiques, mésologiques, météorologiques, intentions expérimentales Des exemples provenant d'échantillonnage en écologie serviront d'illustration.

J2 – 13h30-16h30

Modélisation de données environnementales

Par Franck TORRE, AMU-IMBE & Maxime LOGEZ, INRAE-RiverLy

Après un rappel sur tests d'hypothèses et le modèle linéaire, l'objectif de cette session est d'initier les utilisateurs aux modèles linéaires généralisés, GLM, à travers des exemples pratiques pris soit en sciences médicales soit en sciences environnementales. Très souvent de par la nature de la variable expliquée, l'hypothèse de sa normalité ne peut être envisagée et il convient d'utiliser d'autres outils statistiques que les modèles linéaires classiques. Les GLMs sont des extensions des modèles linéaires à des distributions non normales comme la loi de Poisson ou la loi Binomiale, adaptées à des variables de comptage ou des données de présence-absence (proportions). Pour pouvoir modéliser des variables avec de telles distributions nous aborderons au cours de cette session la régression de Poisson et la régression logistique.

J2 – 17h30-19h30

Présentation de travaux & problématiques quantitatives

Par les participants volontaires

Un temps – en soirée - pourra être réservé à la consultation statistique.

J3 – Mercredi 30 mars 2022

J3 – 9h30-12h30

Notions avancées sur R

Par Maxime LOGEZ, INRAE-RiverLy

Ce module a pour but de familiariser les utilisateurs avec la programmation en R, avec d'une part l'usage et la création de fonctions, l'utilisation d'outils de programmations classiques et très utilisés que sont les boucles et leurs pendants (fonctions de la famille des apply) ainsi que les différents éléments de langages indispensables. Nous montrerons les possibilités du logiciel en termes de lecture de données (lecture conditionnelle de tableau, ...) ainsi que sur l'utilisation de représentations graphiques interactives.

J3 – 9h30-12h30

Après-midi récréative en Ubaye

Une activité culturelle ou touristique sera programmée en lien avec des professionnels de la montagne

<https://www.sud-alpes-aventures.fr/>



Pont Chatelet – Vallée d'Ubaye

J3 – 17h30-19h30

Graphiques ggplot

Par Maxime LOGEZ, INRAE-RiverLy

La librairie ggplot2 offre de très nombreuses possibilités de représentation graphiques simples (nuages de points, histogramme, courbe de densité, ...) et complexes (multi-panneau). Elle s'intègre pleinement dans l'univers « tidyverse ». Le but de cette session sera d'utiliser les fonctions de mise en forme des tableaux des librairies *dplyr* et *tidyr* pour ensuite réaliser des représentations graphiques avec ggplot2 et les customiser.

Un temps – en soirée - pourra être réservé à la consultation statistique.

J4 – Jeudi 31 mars 2022

J4 – 9h30-12h30

Analyse spatiale 1/2

Par Alberte Bondeau, CNRS-IMBE

Après une présentation rapide des types de questions et de données auxquelles s'appliquent les méthodes des statistiques spatiales (processus ponctuels, analyses sur réseaux et sur grille, géostatistique), la demi-journée sera consacrée à une introduction des concepts et méthodes de la Géostatistique au travers d'exemples et de petits programmes sous R.

J4 – 13h30-16h30

Analyse spatiale 2/2

Par Alberte Bondeau, CNRS-IMBE

Dans la continuité de la première demi-journée :

Visualisation et description de données spatiales. Hypothèses générales et modèles utilisés en géostatistique (utilisation de méthodes de simulations pour visualiser le potentiel et les limites du cadre théorique). Outils d'analyse de la variabilité spatiale : variogramme expérimental, fonction de covariance spatiale, choix de modèles et ajustement (présentation autour d'exemples). Méthodes d'interpolation par Krigeage (ordinaire et universel) dans des cas simples et univariés. Influence du choix du modèle et réflexion sur les types d'échantillonnage.

J4 – 17h30-19h30

Présentation de travaux & problématiques quantitatives

Par les participants volontaires

Un temps – en soirée - pourra être réservé à la consultation statistique.

J5 – Vendredi 1^{er} avril 2022

J5 – 9h30-12h30

Retour d'expérience et application à un cas d'étude sur le changement climatique,

Analyses statistiques et rédaction de rapport dynamique sous R (notebook)

Par Joël GUIOT, CNRS-Cerege émérite

Le cas d'étude consiste à étudier les tendances des évolutions des températures depuis 1900 dans différentes parties du monde et à la comparer à la tendance globale. Ce cas permettra d'utiliser des fonctions de régressions, d'analyse spatiale et des graphiques sur ggplot2, mais le point le plus important consistera à utiliser Rstudio pour écrire des rapports basés sur les résultats des analyses grâce à son notebook. C'est une facilité très utile pour partager les résultats avec d'autres, mais également pour introduire de la flexibilité dans les analyses.

J5 – après-midi

Retour à Marseille

Affiliations des conférenciers

