



Journées d'analyse statistique des données sur R

Edition 2020

CEREGE, 2^e étage porte ouest, salle 301 (géomatique)
16, 17, 18 & 20 mars 2020

Inscriptions : Joëlle Cavalieri, FR ECCOREV, tel 0442971521 - mel : cavalieri@cerege.fr
La formation est ouverte en priorité aux doctorants et jeunes chercheurs des UMR affiliées à ECCOREV.

Ces journées se dérouleront sur ordinateur, avec le logiciel R. 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. On dispose au CEREGE de 20 ordinateurs en réseau WIFI. La salle étant relativement petite, on ne pourra accueillir qu'une trentaine d'étudiants par module. Le public visé sera en priorité: les thésards, post-docs et jeunes chercheurs des institutions membres d'ECCOREV, selon la place disponible. L'objectif est de rendre l'utilisateur suffisamment autonome dans les domaines les plus « populaires » de la statistique, de manière à ce qu'il soit capable de pratiquer directement les cas standards et d'être capable de personnaliser son approche.

J1 – Lundi 16 mars 2020 – Introduction & Initiation à la syntaxe de R

J1 – 9h30-12h30

Introduction à R

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

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.

J1 – 13h30-16h30

Analyse de données environnementales multivariées 1/2

Par Franck TORRE, IMBE-AMU

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. Des exemples provenant d'échantillonnage en écologie serviront d'illustration.

J2 – Mardi 17 mars 2020 – Analyse multivariée en environnement

J2 – 9h30-12h30

Analyse de données environnementales multivariées 2/2

Par Franck TORRE, IMBE-AMU

Les méthodes de couplage de tableaux permettent d'étudier le lien entre deux tableaux. On présentera l'analyse de coïnertie qui permet d'étudier la structure commune à deux tableaux contenant différents descripteurs sur les mêmes individus. On présentera également les analyses multivariées explicatives type analyse de redondances (RDA) ou analyse des correspondances sous contrainte (CCA). 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

J2 – 13h30-16h30

Notions avancées sur R

Par Maxime LOGEZ, INRAE

Ce module s'adresse à des utilisateurs de R ayant déjà pratiqués les bases du logiciel. Au travers d'exemples pris en sciences de l'environnement, nous montrerons l'utilisation de certains packages ainsi que la création de fonctions sous R. 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. Toutes ces notions seront abordées en utilisant des méthodes statistiques classiques (régression multiple, ACP, ...).

J3 – Mercredi 18 mars – Graphiques élaborés & Analyse spatiale

J1 – 9h30-12h30

Graphiques ggplot

Par Maxime LOGEZ, INRAE

xxx

J3 – 13h30-16h30

Analyse spatiale 1/2

Par Alberte Bondeau, CNRS-IMBE

La demi-journée commencera par une présentation des types de questions et de données géolocalisées (données surfaciques, données ponctuelles, données continues) auxquelles s'appliquent les outils cartographiques et les méthodes d'analyse spatiales (analyse spatiale descriptive, géomatique, géostatistique). A l'aide de petits programmes sous R, nous aborderons la combinaison de données spatiales disponibles sous différents formats (vecteur/raster) et différentes projections, typiques des données habituellement utilisées dans les géosciences, leur visualisation cartographique et les analyses statistiques spatiales ou spatio-temporelles possibles.

J4 matin – Vendredi 20 mars – Analyse spatiale,

J4 – 9h30-12h30

Analyse spatiale 2/2

Par Alberte Bondeau, CNRS-IMBE

Lors de cette demi-journée, nous aborderons la modélisation géostatistique. Nous étudierons en particulier les outils d'analyse de la variabilité spatiale et le choix des modèles pour représenter un variogramme. Nous appliquerons les méthodes d'interpolation spatiale par Krigeage à des exemples représentatifs de travaux conduits en géosciences.

J4 – 13h30-16h30

Modèle linéaire généralisé

Par Maxime LOGEZ, INRAE

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. Les modèles linéaires classiques, tels que la régression linéaire, l'ANOVA ou encore l'ANCOVA font l'hypothèse que la variable à expliquer (Y) suit une loi normale et que la moyenne de Y (l'espérance) varie selon une équation de la forme $Y = a_i X_i + b$ avec X_i la ou les variable(s) explicative(s). 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.