DellaDATA
Claire DELLA - VEDOVA

Biostatistiques appliquées – Niveau 2

Modélisation statistique et analyses multivariées avec R (3 jours – 21 h)

A qui s'adresse cette formation ?

Cette formation s'adresse aux personnes maîtrisant les fondamentaux des biostatistiques et souhaitant approfondir leurs compétences pour réaliser ou interpréter des analyses statistiques plus avancées.

Elle est particulièrement adaptée aux :

- Professionnel(le)s de la santé, de l'environnement et de la recherche.
- Chercheur(e)s, enseignant(e)s-chercheur(e)s, doctorant(e)s et post-doctorant(e)s.
- Ingénieur(e)s, chargé(e)s d'études et professionnel(le)s amené(e)s à analyser des données scientifiques.
- Personnes utilisant R pour leurs analyses et souhaitant aller au-delà des tests statistiques usuels.

Cette formation constitue la suite logique du niveau 1 « Fondamentaux et tests statistiques usuels avec R », sans qu'il soit nécessaire d'avoir suivi cette première formation si les prérequis sont maîtrisés.

Ce que cette formation n'est pas

Cette formation n'a pas vocation à :

- Présenter de manière exhaustive l'ensemble des méthodes de modélisation et d'analyse multivariée en trois jours.
- Se substituer à un enseignement universitaire approfondi de la modélisation statistique.
- Proposer des procédures automatiques sans réflexion sur la pertinence des méthodes et leurs conditions d'application.

Elle vise au contraire à développer une compréhension pratique des principales méthodes, à apprendre à choisir des analyses adaptées aux questions scientifiques et à interpréter leurs résultats avec rigueur.

Prérequis

Les participants doivent maîtriser les notions fondamentales de statistique descriptive et inférentielle, notamment :

- Les principaux tests de comparaison de moyennes et de proportions.
- L'ANOVA à un facteur.
- La corrélation et la régression linéaire simple.
- L'interprétation des valeurs p et des intervalles de confiance.
- Les principes généraux des tests statistiques et leurs conditions d'application.

Une connaissance des bases de R est également nécessaire : importation de données, manipulation d'objets simples, exécution de scripts et compréhension de code R existant.

Il n'est pas nécessaire de maîtriser les modèles statistiques abordés pendant la formation.

Objectifs et compétences visées

Cette formation vise à permettre aux participant(e)s de comprendre, choisir, mettre en œuvre et interpréter les principales méthodes de modélisation statistique et d'analyse multivariée utilisées en recherche.

À l'issue de la formation, les participant(e)s seront capables de :

- Identifier les situations nécessitant une modélisation statistique et comprendre les principes de l'ajustement.
- Choisir un modèle adapté à la nature de la variable réponse et à la question scientifique.
- Réaliser et interpréter une ANOVA factorielle et une ANCOVA.
- Construire et interpréter des modèles de régression linéaire multiple, logistique et de Poisson.
- Vérifier les principales hypothèses et conditions d'application des modèles.
- Évaluer les performances discriminantes d'un modèle logistique à l'aide d'une courbe ROC.
- Réaliser et interpréter une analyse de survie avec Kaplan-Meier et le modèle de Cox.
- Analyser des données répétées et identifier les situations nécessitant des modèles mixtes.
- Réaliser et interpréter une analyse en composantes principales (ACP).
- Mettre en œuvre une classification ascendante hiérarchique (CAH) et interpréter les groupes obtenus.

Ces compétences permettront aux participants de mieux justifier leurs choix méthodologiques, d'analyser des données complexes et de présenter des résultats statistiques pertinents dans un contexte scientifique ou professionnel.

Ces compétences permettent de sécuriser les analyses statistiques, de mieux justifier les choix méthodologiques et de produire des résultats interprétables et exploitables dans un cadre professionnel.

Programme de la formation

Module 1 – Comprendre les principes de la modélisation statistique

Objectif : comprendre l'intérêt de la modélisation pour étudier simultanément plusieurs variables et prendre en compte les facteurs de confusion.

- Limites des analyses bivariées.
- Variables réponses et variables explicatives.
- Variables de confusion et ajustement statistique.
- Principes généraux des modèles statistiques.
- Choix d'un modèle selon la nature de la variable réponse.

Module 2 – ANOVA factorielle et ANCOVA

Objectif : analyser l'effet de plusieurs facteurs sur une variable quantitative et prendre en compte une covariable.

- ANOVA à deux facteurs.
- Effets principaux et interactions.
- Analyse de covariance (ANCOVA).
- Ajustement sur une covariable quantitative.
- Interprétation des résultats et représentations graphiques.
- Applications pratiques sous R.

Module 3 – Régression linéaire multiple

Objectif : construire et interpréter un modèle permettant d'étudier l'association entre une variable quantitative et plusieurs variables explicatives.

- Construction d'un modèle de régression linéaire multiple.
- Variables explicatives quantitatives et qualitatives.
- Variables indicatrices et catégories de référence.
- Interprétation des coefficients et des intervalles de confiance.
- Prise en compte des interactions.
- Vérification des hypothèses et diagnostic du modèle.
- Applications pratiques sous R.

Module 4 – Régression logistique et courbes ROC

Objectif : modéliser la probabilité de survenue d'un événement binaire et évaluer les performances discriminantes du modèle.

- Principes de la régression logistique.
- Construction d'un modèle de régression logistique multiple.
- Odds ratios bruts et ajustés.
- Interprétation des coefficients et des probabilités prédites.
- Utilisation des probabilités prédites pour la classification.
- Sensibilité, spécificité et choix d'un seuil de classification.
- Courbes ROC et aire sous la courbe (AUC).
- Applications pratiques sous R.

Module 5 – Régression de Poisson

Objectif : comprendre et mettre en œuvre un modèle adapté à l'analyse des données de comptage et des taux d'incidence.

- Principes de la régression de Poisson.
- Construction et interprétation du modèle.
- Rapports de taux et intervalles de confiance.
- Prise en compte du temps d'exposition.
- Notion de surdispersion et limites du modèle.
- Applications pratiques sous R.

Module 6 – Analyse de survie

Objectif : analyser le délai de survenue d'un événement et étudier l'association entre plusieurs facteurs et le risque de survenue.

- Particularités des données de survie et notion de censure.
- Estimation et interprétation des courbes de Kaplan-Meier.
- Comparaison des courbes de survie par le test du log-rank.
- Principes du modèle de régression de Cox.
- Interprétation des hazard ratios bruts et ajustés.
- Introduction à l'hypothèse des risques proportionnels.
- Applications pratiques sous R.

Module 7 – Analyser des données répétées

Objectif : comprendre les particularités des données longitudinales et choisir une approche adaptée à la dépendance entre observations.

- Particularités des données répétées et longitudinales.
- ANOVA à mesures répétées.
- Vérification des hypothèses, notamment la sphéricité.
- Alternatives non paramétriques.
- Introduction aux modèles mixtes et à leurs situations d'utilisation.
- Applications pratiques sous R.

Module 8 – Analyse en composantes principales (ACP)

Objectif : explorer des données multidimensionnelles, réduire le nombre de dimensions et interpréter les principales structures observées.

- Objectifs et principes de l'ACP.
- Réduction de dimension et synthèse de l'information.
- Valeurs propres, variance expliquée et choix des axes.
- Cercle des corrélations et représentation des variables.
- Projection des individus sur les axes factoriels.
- Contributions et qualité de représentation.
- Interprétation des résultats.
- Applications pratiques sous R.

Module 9 – Classification et clustering

Objectif : identifier des groupes d'individus présentant des caractéristiques similaires et interpréter les classifications obtenues.

- Principes de la classification non supervisée.
- Notions de distance et de similarité entre individus.
- Classification ascendante hiérarchique (CAH).
- Construction et interprétation d'un dendrogramme.
- Choix du nombre de classes.
- Caractérisation des groupes obtenus.
- Complémentarité entre ACP et classification.
- Applications pratiques sous R.

Modalités pratiques

⏱️ Durée : 3 jours (21 heures)

💻 Format : distanciel (visioconférence), présentiel possible selon les besoins

👥 Effectif : groupes à taille limitée (environ 5 participant(e)s)

📘 Pédagogie et organisation :

Alternance d'apports théoriques et de mises en pratique
Supports remis aux participant(e)s. Temps dédiés aux questions et à l'interprétation

🧰 Logiciels : R et R Studio

📝 Évaluation des acquis : Exercices pratiques, mises en situation et échanges oraux tout au long de la formation.

🏷️ Tarif

1 650 € HT par participant(e) pour la formation de 3 jours (21 h) - Supports pédagogiques inclus.

Les tarifs en intra-entreprise sont établis sur devis, en fonction du format et des modalités de la formation.
Organisme exonéré de TVA en vertu de l'article 261-4-4°a du CGI.

✓ Suivi et assiduité : Émargement par demi-journée.

Adaptation et accessibilité :

Les exemples peuvent être adaptés au contexte des participants en intra-entreprise.
Toute situation de handicap ou besoin spécifique peut être signalé en amont afin d'étudier les adaptations possibles

Informations de contact

Organisme de formation : **DellaDATA**
NDA : 93040100204 auprès du préfet de région PACA
SIRET : 90838795400021
Organisme exonéré de TVA en vertu de l'article 261-4-4°a du CGI
Adresse : 80. Bd Charles de Gaulle - BâtD - 04100 Manosque
Formatrice et interlocutrice : Claire Della Vedova (claire@delladata.fr / 06.64.98.21.42)