

# Formation Réaliser ses analyses statistiques avec R

**Objectifs :** Apprendre à utiliser le logiciel R pour analyser des données. Mettre en oeuvre dans R les méthodes de statistique descriptive, décisionnelle, analyse de la variance, régression linéaire et analyse de données multidimensionnelles.

**Compétences visées :** - Importer, manipuler et analyser des données dans R

- Effectuer des analyses statistiques univariées
- Établir avec R des intervalles de confiance autour de paramètres comme moyenne, proportion et variance
- Réaliser les principaux tests d'hypothèses paramétriques et non paramétriques dans le logiciel
- Étudier numériquement et graphiquement la liaison entre 2 variables quantitatives, qualitatives ou de chaque nature
- Mettre en oeuvre des analyses de la variance à un ou plusieurs facteurs pour la comparaison de moyennes de population
- Élaborer un modèle établissant une relation linéaire entre une variable à expliquer et une (régression simple) ou plusieurs (régression multiple) variables explicatives.
- Découvrir les principales méthodes d'analyse multidimensionnelle des données à l'aide du package FactoMineR

**Durée :** 3 jour(s) (21 heures)

**Public :** Toute personne souhaitant analyser des données avec R

**Pré-requis :** Pour suivre ce stage dans de bonnes conditions, il est recommandé d'avoir suivi en amont la formation [R - Prise en main, analyses statistiques et graphiques](#)

**Méthode pédagogique :** Pédagogie active mêlant exposés, exercices et applications pratiques dans le logiciel R.

**Modalités d'évaluation :** Un formulaire d'auto-évaluation proposé en amont de la formation nous permettra d'évaluer votre niveau et de recueillir vos attentes. Ce même formulaire soumis en aval de la formation fournira une appréciation de votre progression.

Des exercices pratiques seront proposés à la fin de chaque séquence pédagogique pour l'évaluation des acquis.

En fin de formation, vous serez amené(e) à renseigner un questionnaire d'évaluation à chaud.

Une attestation de formation vous sera adressée à l'issue de la session.

Trois mois après votre formation, vous recevrez par email un formulaire d'évaluation à froid sur l'utilisation des acquis de la formation.

**Accessibilité :** Vous souhaitez suivre notre formation Formation par logiciel et êtes en situation de handicap ? Merci de nous contacter afin que nous puissions envisager les adaptations nécessaires et vous garantir de bonnes conditions d'apprentissage

**Tarifs :**

- Présentiel : 1650 € HT
- Distanciel : 1500 € HT
- (-10% pour 2 inscrits, -20% dès 3 inscrits)

**Option(s) :**

- Forfait déjeuners : 60 € HT

**Nos prochaines sessions****Bordeaux**

du 25 au 27 juin 2025

**Distance**

du 12 au 14 mars 2025

du 18 au 20 juin 2025

du 12 au 14 novembre 2025

**Lyon**

du 1 au 3 avril 2025

du 1 au 3 octobre 2025

**Paris**

du 12 au 14 mars 2025

du 4 au 6 juin 2025

du 3 au 5 septembre 2025

du 10 au 12 décembre 2025

**Toulouse**

du 14 au 16 mai 2025

du 15 au 17 octobre 2025

**Programme :**

## - Statistiques descriptives

- Gestion d'un jeu de données - dataframe  
Importation, caractérisation, sélection, sous-ensembles
- Premières analyses d'un jeu de données  
Premières vérifications, valeurs manquantes, recodage
- Résumé d'une variable quantitative - numeric  
Indicateurs numériques, représentations graphiques
- Résumé d'une variable qualitative - factor  
Indicateurs numériques, représentations graphiques

## - Intervalle de confiance

- Le raisonnement à partir d'un échantillon  
Généralités, échantillonnage, estimation d'un paramètre
- Intervalle de confiance d'une moyenne
- Intervalle de confiance d'une proportion
- Intervalle de confiance d'une variance

## - Tests d'hypothèses

- Qu'est-ce qu'un test d'hypothèse ?  
Généralités, règle de décision, risques d'erreur, puissance
- Les tests de conformité ou de comparaison à une norme  
Conformité d'une moyenne, d'une proportion
- Les tests de comparaison de deux populations  
Comparaison de deux moyennes, de deux proportions
- Test d'ajustement à une loi de probabilité normale  
Le test de Shapiro-Wilk
- Introduction aux tests non paramétriques  
Test de Wilcoxon, Mann et Whitney, Kruskal Wallis, Friedman

## - Liaisons entre deux variables

- Liaison entre deux variables quantitatives  
Nuage de points, intensité de la liaison, significativité
- Liaison entre deux variables qualitatives  
Tableau de contingence, intensité et significativité du lien de dépendance : test du khi2
- Liaison entre une variable qualitative et quantitative  
Comparaison de plusieurs populations, le rapport de corrélation
- Liaisons entre plusieurs variables  
Approches graphiques : matrice de nuages de points, treillis  
Caractériser des sous-populations par plusieurs variables

## - L'analyse de la variance - Anova

- Analyse de la variance à un facteur  
Variabilité inter, intra, totale - Rapport de corrélation - Le test de Fisher
- Comparaisons multiples de moyennes  
Analyses post hoc, la procédure de Tukey

- Analyse de la variance à deux facteurs et interaction
- Extensions de l'Anova  
Modèle à effet fixe ou aléatoire, modèle hiérarchisé

## - Régression linéaire simple et multiple

- De la corrélation à la régression  
L'intérêt d'un modèle - Variables à expliquer, explicatives, erreur
- La régression linéaire simple  
Ajustement par la méthode des moindres carrés - Tests et validation du modèle
- La régression linéaire multiple
- Choix d'un modèle de régression  
Pourquoi sélectionner un sous-ensemble de variables explicatives ?  
Les différentes approches et critères de sélection d'un modèle

## - Analyse de données multidimensionnelles

- Un panorama des méthodes multidimensionnelles  
Analyses factorielles, classification - Le package FactoMineR
- ACP : Analyse en Composantes principales
- AFC : Analyse Factorielle des Correspondances
- ACM : Analyse des correspondances Multiples
- CAH : Classification Ascendante Hiérarchique

*Date de dernière modification : 5 novembre 2024*