

# Les fondamentaux de l'analyse statistique avec R

## INFORMATIONS GÉNÉRALES

**Type de formation :** Formation continue

**Éligible au CPF :** Non

**Domaine :** IA, Big Data et Bases de données

**Action collective :** Non

**Filière :** Big Data

**Rubrique :** ElastickSearch,...

**Code de formation :** BD0103

## € Tarifs

**Prix public :** 2195 €

### Tarif & financement :

Nous vous accompagnons pour trouver la meilleure solution de financement parmi les suivantes :

- **Le plan de développement des compétences de votre entreprise** (rapprochez-vous de votre service RH).
- **Le dispositif FNE-Formation.**
- **L'OPCO** (opérateurs de compétences) de votre entreprise.
- **Pôle Emploi** sous réserve de l'acceptation de votre dossier par votre conseiller Pôle Emploi.
- **CPF -MonCompteFormation.**

Contactez nous pour plus d'informations.

## PRÉSENTATION

### Objectifs & compétences

Savoir installer R  
Comprendre comment manipuler des données avec R  
Savoir importer et exporter des données  
Être en mesure de réaliser des analyses statistiques basiques avec R  
Savoir restituer des résultats à l'aide de graphiques

### Public visé

Ingénieurs  
Analystes  
Data analysts  
Toute personne intéressée par l'analyse statistique avec R

### Pré-requis

Avoir suivi la formation "Les fondamentaux de la statistique appliquée" (BI090) ou connaissances mathématiques équivalentes  
Être familier avec l'environnement Microsoft Windows

## 📍 Lieux & Horaires

**Campus :** Ensemble des sites

**Durée :** 21 heures

### Délai d'accès :

Jusqu'à 8 jours avant le début de la formation.

**Distanciel possible :** Oui

## PROGRAMME

### 1 - Introduction

Qu'est-ce que R ?  
Avantages et inconvénients  
Solutions concurrentes gratuites ou payantes

### 2 - Installation

Installation de R ou Microsoft R Open sur MS Windows ou Scientific Linux  
Découverte de l'environnement  
Utiliser l'historique des commandes  
Exemple d'environnement superposé (R-Studio)  
Comment citer le logiciel dans une publication scientifique  
Comment citer des packages dans une publication scientifique

### 3 - Utilisation

Vider la console de commande  
Utilisation de l'aide  
Changer la langue de l'interface  
Quitter en ligne de commande  
Changer le dossier de travail par défaut temporairement ou à chaque session (\*.Rprofile)  
Changer définitivement le dossier par défaut des packages (\*.Rprofile)  
Sauver/Charger l'espace de travail (\*.Rdata)  
Sauver/Charger/Exécuter un historique des commandes (\*.Rhistory)  
Sauver les commandes et sorties dans un fichier (\*.txt)  
Sauver/Charger un script (\*.R)

### 4 - Manipulation de packages

Installer/Désinstaller/Mettre à jour des packages  
Obtenir des informations systèmes sur les packages  
Écrire plusieurs commandes sur une ligne  
Ajouter des commentaires

## 📅 Prochaines sessions

Cliquez sur la date choisie pour vous inscrire :

#### ■ 23 / 09 / 2024

📍 : Ensemble des sites  
✓ : Distanciel possible  
🕒 : 21 heures  
📅 : 3 jours

#### ■ 04 / 11 / 2024

📍 : Ensemble des sites  
✓ : Distanciel possible  
🕒 : 21 heures  
📅 : 3 jours

**5 - Types de données**

Manipulations de scalaires (réels)  
Manipulations de nombres complexes  
Manipulations de variables  
Manipulations de vecteurs  
Manipulations de matrices  
Manipulations de textes  
Manipulations de dates et de durées  
Création/Édition de données en ligne de commande  
Gérer les variables dans la mémoire

**6 - Import et export de données**

Importer/Exporter des données d'Excel  
Importer/Exporter des données en \*.csv et gestion du passage d'encodage  
Linux/Windows  
Importer/Exporter des fichiers SAS  
Importer des données \*.csv du web  
Fusionner des fichiers \*.csv  
Importer de données de MS Access  
Importer des données \*.xml du web  
Importer des données de MySQL  
Importer des données d'Oracle (Express)  
Importer des fichiers \*.json

**7 - Manipulation de données**

Utiliser le SQL  
Quelques Data set  
Manipuler les data frames  
Accélérer l'accès aux colonnes  
Renommer les colonnes  
Appliquer des rangs  
Trier des données  
Filtrer des données  
Réaliser des sous-sélections  
Fusionner des données  
Supprimer les doublons  
Échantillonnage  
Empiler/Déempiler des données

**8 - Analyse de données**

Synthétiser des données (tables de contingence)  
Travailler avec des valeurs absentes  
Définir le nombre de décimales de chaque sortie  
Générer des variables pseudo-aléatoires  
Statistiques descriptives simples (comptage, éléments uniques, moyenne, max, min, centiles, somme, écart-type biaisé/non biaisé, cv, médiane, etc.)  
Plotter (tracer) des fonctions algébriques  
Racines d'équations univariées  
Intégration numérique  
Intégration algébrique  
Dérivation algébrique/numérique  
Optimisation linéaire uni ou multidimensionnelle  
Optimisation sous contrainte  
Programmation linéaire

**MODALITÉS****Modalités**

**Modalités** : en présentiel, distanciel ou mixte – Horaires de 9H à 12H30 et de 14H à 17H30 soit 7H – Intra et Inter entreprise

**Pédagogie** : essentiellement participative et ludique, centrée sur l'expérience, l'immersion et la mise en pratique. Alternance d'apports théoriques et d'outils pratiques.

**Ressources techniques et pédagogiques** : Support de formation au format PDF ou PPT Ordinateur, vidéoprojecteur, Tableau blanc, Visioconférence : Cisco Webex / Teams / Zoom

**Pendant la formation** : mises en situation, autodiagnostic, travail individuel ou en sous-groupe sur des cas réels

**Méthode**

**Fin de formation** : entretien individuel

**Satisfaction des participants** : questionnaire de satisfaction réalisé en fin de

formation

**Assiduité** : certificat de réalisation (validation des acquis)

### **Les plus de la formation**

Une formation très pratique : 90% du temps de la formation est dédié à la mise en pratique pour une meilleure assimilation de notions de base.

Des consultants expérimentés partagent leur savoir-faire avec les participants.