

Plan de formation

LANGAGE R

TRAITEMENT DE DONNEES ET ANALYSES STATISTIQUES

DUREE : 3 jours

VILLES : Nantes, Paris, Montpellier, Lyon, Bordeaux, Lyon, Angers, Brest

Contact Formation

Marie KIEFFER

☎ 02 34 09 31 72

✉ formation@nextdecision.fr

www.next-decision.fr

PRESENTATION DE LA FORMATION

- ▶ Cette formation est une introduction au Langage R dans le but de manipuler et traiter des données, de faire des analyses statistiques et des restitutions graphiques.

OBJECTIFS DE LA FORMATION

- ▶ Maîtriser les bases du logiciel pour le traitement des données et pour l'analyse statistique
- ▶ Permettre à l'utilisateur de créer et modifier des programmes R

PREREQUIS DE LA FORMATION

- ▶ Connaître les principes de l'informatique bureautique (traitements de données avec un outil de type tableau, ex : Excel...)
- ▶ Avoir des connaissances en statistiques est un plus

PUBLIC VISE DE LA FORMATION

- ▶ La formation Langage R traitement de données et analyses statistiques s'adresse à des débutants.

Programme de la formation

▶ PRÉSENTATION DU LOGICIEL R

- Logiciel en Open source
- Les capacités de R
- L'utilisation du logiciel R
- Dans quels domaines ?

▶ PRÉSENTATION DU LOGICIEL R-STUDIO

- La console
- Le script
- L'environnement
- Les files
- Les plots
- Les packages
- Help

▶ PRÉSENTATION DES OBJETS ET DU FONCTIONNEMENT DE R

- Les vecteurs et les facteurs
- Les listes
- Les matrices et les arrays
- Les data frames
- Les fonctions

▶ LES PACKAGES R

- Installation d'un package
- Chargement d'un package
- Les packages R les plus utilisés

▶ MANIPULATION D'OBJETS DANS R

- Création d'objets
 - Vecteur
 - Facteur
 - Liste
 - Matrice
 - Array
 - Data Frame
 - Fonction
 - Suppression d'objets
- Import et export des données
 - Importation / exportation de fichiers plats (CSV, Text, ...)
 - Importation / exportation de fichiers Excel (.xls, xlsx)
 - Importation / exportation de fichiers R (.RData, .R)
- Le traitement des données
 - Trier un jeu de données
 - Renommer des colonnes et des lignes d'un jeu de données
 - Sélectionner des données au sein d'un tableau
 - Supprimer des données d'un jeu de données
 - Utilisation de la fonction merge pour joindre des data frames

La concaténation

- Ajouter des colonnes à un data frame avec la fonction cbind
- Ajouter des lignes à un data frame avec la fonction rbind

Création d'indicateurs

▶ STATISTIQUES DESCRIPTIVES DANS R

Les fonctions usuelles

- Apply
- Aggregate

Les indicateurs de statistiques simples

- Summary
- Sum (Somme)
- Mean (Moyenne)
- Var (Variance)
- Sd (Ecart type)
- Min (Minimum)
- Max (Maximum)

Les représentations de graphiques simples

- Plot (Graphique de base)
- Pie (Diagramme circulaire)
- Barplot (Diagramme en barres)
- Boxplot (Boîtes à moustaches)
- Exemples de graphiques avec le package Ggplot2

▶ ESTIMATION ET TESTS STATISTIQUES

Les intervalles de confiance

- Utilisation de la fonction t.test pour un intervalle de confiance d'une moyenne
- Utilisation de la fonction prop.test pour un intervalle de confiance d'une proportion

Les tests de normalité

- La fonction ks.test (Test de Kolmogorov-Smirnov)
- La fonction shapiro.test (Test de Shapiro-Wilk)

Les tests de comparaison de moyennes et de proportions

- La fonction t.test (Test de Student)
- La fonction wilcox.test (Test de Wilcoxon-Man-Whitney)
- La fonction fisher.test (Test de Fisher)

La fonction chisq.test (Test du χ^2)

▶ LIEN ENTRE PLUSIEURS VARIABLES

Obtenir le lien de corrélation entre deux variables avec la fonction cor

Faire une régression linéaire multiple avec les fonctions lm et (Linear model) et glm (Generalized linear model)

Analyser la variance avec la fonction anova

Procéder à un test sur plusieurs variables avec la fonction kruskal.test

▶ LES ANALYSES MULTIVARIÉES

Réaliser une classification ascendante hiérarchique avec les fonctions dist et hclust

La fonction kmean

Faire une régression logistique avec la fonction glm

► **POUR ALLER PLUS LOIN AVEC R**

- Exemple d'interface avec RShiny
- Exemple d'utilisation de RMarkdown
- Utilisation avec un autre langage
 - SQL
 - Python

Les petits plus de Next Decision

Tarification et modalités de prise en charge.

Nos formateurs sont tous consultants formateurs, nos formations sont facturées à la journée du formateur et non pas à celle du stagiaire. Notre organisme est certifié Datadock, nos formations peuvent donc être prises en charge par votre OPCO.

Les moyens pédagogiques

Les supports pédagogiques sont imprimés par nos soins et transmis à chacun des stagiaires. Ce support est projeté via un vidéo projecteur afin d'animer la formation. Des exercices d'application ou études de cas sont prévus afin de valider les acquis des stagiaires.

Les modalités de suivi de l'exécution

Les formateurs se rendent à la formation avec l'ensemble des documents de fin de session. A la fin de chaque stage, la feuille d'émargement est signée pour la validation de la participation des stagiaires et les évaluations sont distribuées puis complétées par chacun des stagiaires.

Les Horaires de formation

Les sessions de formation débutent à 09h30 et se terminent à 17h00. Pour les stages animés dans vos locaux, les horaires peuvent être aménagés selon vos disponibilités.

Nos particularités

Formation Intra-Entreprise

Nous proposons des formations intra entreprise durant lesquelles nos consultants formateurs pourront adapter le contenu de la formation à votre besoin. Vous pouvez commander une formation catalogue et effectuer cette session dans vos propres locaux. Ainsi, seul le formateur se déplace à la rencontre de vos équipes. En amont des sessions de formation, le consultant formateur s'imprègne du contexte du projet afin d'adapter la session de formation à vos besoins.

Formation Inter-entreprises

Dans nos locaux ou dans ceux de nos partenaires. Ce format présente l'avantage de partager et d'échanger avec d'autres professionnels venus d'autres horizons sur les sujets couverts par le stage, ce qui facilite l'application du contenu de la formation dans le contexte de l'entreprise au quotidien.