

**Plan de cours N° : 1089****Durée :** 4 jours (28h)**Participants**

Développeur, chef de projets proche du développement, ingénieur scientifique sachant coder.

**Pré-Requis**

Maîtriser l'algorithmique, avoir une appétence pour les mathématiques, la connaissance des statistiques est un plus.

**Objectifs**

Savoir mettre en place une stratégie de traitement de données avec le langage R grâce à plusieurs modèles mathématiques.

**Méthode pédagogique**

Alternance entre apports théoriques et exercices pratiques  
Support de cours fourni lors de la formation

**Moyens d'encadrement mis en oeuvre**

1 à 8 personnes maximum par session  
1 poste informatique par personne  
Une assistance post-formation, d'une durée d'un an, sur le contenu de la formation

**Moyens permettant de suivre son exécution et d'en apprécier les résultats**

Emergence par demi-journée  
Evaluation des acquis par mise en situation de travail  
Evaluation qualitative de fin de stage  
Remise d'une attestation individuelle de formation en fin de stage

**Assistance**

formateurs@atp-formation.com

**Introduction à R**

- Installation
- Historique
- RShell
- RStudio
- Les variables
- Les fonctions de base
- Les fonctions algébriques
- Les conversions
- Les fichiers RData
- L'installation de package

**Les Instructions R**

- Les blocs
- Les conditions
- Les boucles
- Les fonctions

**Les vecteurs**

- La création
- Les opérateurs de vecteur
- Les fonctions seq et c
- L'indexation

**Les statistiques avec R**

- Les calculs de distribution
- La loi normale
- Ecart type et variance
- Les fonctions de statistiques autres
- Les quantiles
- La réciprocity des données

**Le graphisme**

- Le graphisme avec plot
- Les différents types de graphiques
- Affichage de plusieurs graphiques

**Les régressions**

- Le concept de régression en statistique
- La régression linéaire
- La fonction lm()
- L'opérateur tilde ~
- Le coefficient de corrélation
- Le loss
- L'erreur quadratique moyenne
- Les régressions polynomiales
- Les régressions fonctionnelles

**Les matrices**

- L'algèbre linéaire en R
- Les matrices
- Le produit et la transposée
- L'inversion de matrice
- Les nombres complexes
- L'algèbre complexe
- La transformée de fourrier FFT

**Les dataframes**

- L'utilité de données tabulaires
- Les dataframes
- L'accès aux dataframes
- L'indexation
- Le filtrage
- Les listes

**L'accès aux sources de données avec R**

- L'accès aux données binaires
- L'accès aux fichiers CSV
- L'accès aux fichiers Excel
- L'accès aux fichiers JSON

**Machine Learning**

- Mise en place d'une machine learning supervisé
- Qu'est-ce qu'un modèle et un dataset ?
- Qu'est-ce qu'une régression ?
- Les différents types de régression
- La régression linéaire
- Gestion du risque et des erreurs
- Quarter d'Ascombe
- Trouver le bon modèle
- La classification
- Filtrage
- Loi normale, variance et écart type
- Apprentissage
- Mesure de la performance
- No Free Lunch
- Mise en place d'une machine learning de régression en R
- Présentation du k-NN