

FORMATION Statistiques
Logiciel R Perfectionnement :
Mise en œuvre des analyses statistiques
Formation en distanciel

3 jours : du **14 au 16 Octobre 2024**

DATE LIMITE D'INSCRIPTION : 09 septembre 2024

Objectifs	<p>A l'issue de cette formation, l'apprenant sera capable de :</p> <ul style="list-style-type: none">• Obtenir de l'aide en ligne, rechercher des commandes permettant de réaliser l'action souhaitée• Manipuler efficacement les facteurs et les data frames• Filtrer un tableau de données• Créer des tableaux de synthèses (tableau des moyennes, des écart-types... par groupes)• Lire et écrire des scripts « avancés » d'analyse des données• Mettre en œuvre et extraire les résultats des méthodes statistiques suivantes :<ul style="list-style-type: none">✓ Analyse descriptive des données✓ Tests paramétriques et non paramétriques classiques : test de Student, test de Wilcoxon✓ ANOVA à 1 facteur✓ ANOVA multi-facteurs✓ Tests post-Hoc de Tukey, de Bonferroni✓ Régression linéaire✓ Analyses factorielles multivariées (ACP, AFC) et classification (CAH)
Public	<p>Cette formation s'adresse à des personnes souhaitant se perfectionner dans l'utilisation du logiciel R dans un contexte de mise en œuvre des analyses statistiques des données. Il ne s'agit donc pas d'une formation sur le fond statistique mais bien d'une formation orientée logiciel. En outre, cette formation n'est pas adaptée aux apprenants à la recherche d'un savoir-faire sur la programmation avancée sous R (boucles, tests conditionnels, création de fonctions d'automatisation performantes...)</p>
Prérequis	<p>Il est indispensable que les participants aient des bonnes connaissances du logiciel R : Importation des données, création et exécution de scripts simples, création et manipulation vecteurs et data frames... quotidiens + analyses stat</p> <p><u>Matériel et environnement</u> :</p> <ul style="list-style-type: none">- Un ordinateur avec micro (+webcam recommandée) et accès à internet- Disposer d'un second écran est conseillé- Pouvoir s'isoler dans un bureau/une salle pendant les temps de formation. <p><u>Outils logiciel</u> :</p> <p>Cette formation n'est pas strictement dédiée à un logiciel. Les exercices et les illustrations se feront à l'aide d'un logiciel "spécialisé" pour ce type d'analyses : XLSTAT, JMP ou R. Le logiciel retenu sera choisi en accord avec celui utilisé par le plus grand nombre des stagiaires.</p>
Durée	3 journées indissociables.
Programme	<ul style="list-style-type: none">• Revoir les fondamentaux sur le logiciel R• Structurer et extraire de l'information d'un tableau de données• Synthétiser l'information d'un tableau de données• Mettre en œuvre des analyses statistiques de bases sous R• Mettre en œuvre une ANOVA sous R• Mettre en œuvre une régression linéaire sous R• Mettre en œuvre une analyse factorielle multivariée et une classification sous R
Méthodes pédagogiques	Explications théoriques suivies de pratiques guidées

Informations pratiques	Formation en distanciel Inscription via le formulaire en ligne à l'adresse https://www.sirene.inserm.fr/
Intervenants	Société SEENOVATE
Contact	Maud Madrid – Chargée de développement des Ressources Humaines Tél. 04 72 13 88 20 Mail : maud.madrid@inserm.fr



SE PERFECTIONNER DANS L'UTILISATION DU LOGICIEL R DANS UN CONTEXTE DE MISE EN ŒUVRE D'ANALYSES STATISTIQUES

NIVEAU : INITIÉ
FILIERE : STAT



DURÉE : 3 JOURS

COURS THÉORIQUES
+ EXERCICES
D'APPLICATIONS

EVALUATION



PREREQUIS :
CONNAISSANCES DE BASES SUR LE LOGICIEL R :
IMPORTATION DES DONNÉES, CREATION ET EXECUTION DE SCRIPTS SIMPLES, CREATION ET MANIPULATION VECTEURS ET DATA FRAMES... QUOTIDIENS + ANALYSES STAT



Pourquoi suivre cette formation ?

Cette formation s'adresse à des personnes souhaitant se perfectionner dans l'utilisation du logiciel R dans un contexte de mise en œuvre des analyses statistiques des données. Il ne s'agit donc pas d'une formation sur le fond statistique mais bien d'une formation orientée logiciel. En outre, cette formation n'est pas adaptée aux apprenants à la recherche d'un savoir-faire sur la programmation avancée sous R (boucles, tests conditionnels, création de fonctions d'automatisation performantes...)

La formation conviendra à des utilisateurs de R venant chercher du savoir-faire logiciel sur la mise en œuvre des méthodes statistiques suivantes :

- ❖ Analyse descriptive des données
- ❖ Tests paramétriques et non paramétriques classiques : test de Student, test de Wilcoxon
- ❖ ANOVA à 1 facteur
- ❖ ANOVA multi-facteurs
- ❖ Tests post-Hoc de Tukey, de Bonferroni
- ❖ Régression linéaire
- ❖ Analyses factorielles multivariées (ACP, AFC) et classification (CAH)

Nom Client

Un temps de la formation (environ 1 journée) sera consacré à la manipulation des données sous R : filtre des données, création d'un tableau de synthèse...

En plus des connaissances de bases sur le logiciel, l'apprenant doit avoir de bonnes connaissances sur les différentes analyses statistiques mises en œuvre au cours de cette formation.

Thèmes principaux

- ❖ Filtre des données
- ❖ Manipulation des vecteurs, facteurs et data frames
- ❖ Création de script pour l'analyse des données
- ❖ Mise en œuvre sous R des méthodes statistiques suivantes :
 - ✓ Analyse descriptive des données
 - ✓ Tests paramétriques et non paramétriques classiques : test de Student, test de Wilcoxon
 - ✓ ANOVA à 1 facteur
 - ✓ ANOVA multi-facteurs
 - ✓ Tests post-Hoc de Tukey, de Bonferroni
 - ✓ Régression linéaire
 - ✓ Analyses factorielles multivariées (ACP, AFC) et classification (CAH)



Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

- ❖ Obtenir de l'aide en ligne, rechercher des commandes permettant de réaliser l'action souhaitée
- ❖ Manipuler efficacement les facteurs et les data frames
- ❖ Filtrer un tableau de données
- ❖ Créer des tableaux de synthèses (tableau des moyennes, des écart-types... par groupes)
- ❖ Lire et écrire des scripts « avancés » d'analyse des données
- ❖ Mettre en œuvre et extraire les résultats des méthodes statistiques suivantes :
 - ✓ Analyse descriptive des données
 - ✓ Tests paramétriques et non paramétriques classiques : test de Student, test de Wilcoxon
 - ✓ ANOVA à 1 facteur
 - ✓ ANOVA multi-facteurs
 - ✓ Tests post-Hoc de Tukey, de Bonferroni
 - ✓ Régression linéaire
 - ✓ Analyses factorielles multivariées (ACP, AFC) et classification (CAH)



Contenu de la formation

Revoir les fondamentaux sur le logiciel R

- ❖ Fonctionnement général du logiciel et documentation en ligne

- ❖ Conseils sur l'organisation d'un projet R
- ❖ Importation des données contenues dans un classeur Excel
- ❖ Installation et chargement d'un package
- ❖ Trucs et astuces

Structurer et extraire de l'information d'un tableau de données

- ❖ Création et manipulation efficace des variables qualitatives (facteurs) et quantitatives
- ❖ Ajout ou suppression d'une variable à un tableau de données
- ❖ Empilement / dépilement d'un tableau de données
- ❖ Filtre d'un tableau de données
 - ✓ A partir des indices des éléments
 - ✓ Par le biais de conditions

Synthétiser l'information d'un tableau de données

- ❖ Commande tapply
- ❖ Commande aggregate
- ❖ Création de tableaux de synthèse (tableaux croisés)

Mettre en œuvre des analyses statistiques de bases sous R

- ❖ Analyse descriptive des données
- ❖ Intervalles de confiance
- ❖ Tests d'hypothèse paramétriques et non paramétriques
 - ✓ Comparaison des moyennes (Student)
 - ✓ Comparaison des médianes (Wilcoxon, Mann-Whitney)
 - ✓ Test de Normalité (Shapiro-Wilk) et QQ-plot

Mettre en œuvre une ANOVA sous R

- ❖ Les commandes aov() et lm()
- ❖ Mise en œuvre d'une ANOVA à un facteur
 - ✓ Tableau d'analyse de la variance
 - ✓ Vérification des conditions d'utilisation du modèle : Normalité et homogénéité des dispersions (test de Bartlett, test de Levene)
 - ✓ Représentation graphique des résultats
 - ✓ Tests post-Hoc (comparaisons multiples) : méthode de Tukey et correction Bonferroni
 - ✓ Tests non paramétriques : Kruskal-Wallis
- ❖ Mise en œuvre d'une ANOVA multi-facteurs
 - ✓ Effets simples, effets croisés
 - ✓ Tableau d'analyse de la variance
 - ✓ Tests post-Hoc
 - ✓ Graphique des interactions
 - ✓ Somme des carrés de type I / somme des carrés de type III

Mettre en œuvre une régression linéaire sous R

- ❖ Ecriture du modèle avec la commande lm()
- ❖ Commandes summary() et anova()
- ❖ Régression linéaire simple
- ❖ Régression linéaire multiple
- ❖ Tableau d'analyse de la variance et estimation des coefficients du modèle
- ❖ Vérification des conditions d'utilisation d'une régression linéaire

Nom Client

- ✓ Normalité des résidus
- ✓ Homogénéité des résidus
- ❖ Diagnostic des multicollinéarités
- ❖ Régression linéaire pas à pas

Mettre en œuvre une analyse factorielle multivariée et une classification sous R

- ❖ Présentation des principaux package
- ❖ ACP (Analyse en Composantes Principales)
 - ✓ Mise en œuvre sous R
 - ✓ Valeurs propres et diagramme des valeurs propres
 - ✓ Cercle factoriel et graphique des individus
 - ✓ Contributions et \cosinus^2
- ❖ AFC (Analyse Factorielle des Correspondances)
 - ✓ Mise en œuvre sous R
 - ✓ Valeurs propres et diagramme des valeurs propres
 - ✓ Graphique des modalités
 - ✓ Contributions et \cosinus^2
- ❖ CAH (Classification Ascendante Hiérarchique)
 - ✓ Commande `hcpc()`
 - ✓ Dendrogramme
 - ✓ Récupération des classes



Notre méthodologie

Les formateurs de **Seenovate** sont des consultants expérimentés et certifiés. Ils appuient leur pédagogie sur leur expertise terrain en conseil et en réalisation de projet en statistique et data science.