

Logiciel R - niveau 2

Orienté analyses statistiques - 3 jours (*)

Référence : STA-N2-R-Stat

Durée : 3 jours soit 21 heures

(*) : La durée proposée est une durée standard. Elle peut être adaptée selon les besoins, les acquis, les niveaux des participants et le contexte.

Filière : Statistiques

Population visée :

Employé – Technicien – Cadre – Chercheur – Etudiant

Public concerné :

Toute personne souhaitant se perfectionner dans l'utilisation du logiciel R dans un contexte de mise en œuvre d'analyses statistiques des données

EN QUELQUES MOTS

Cette formation s'adresse à des personnes souhaitant se perfectionner dans l'utilisation du logiciel R dans un contexte de mise en œuvre d'analyses statistiques des données. Elle conviendra à des utilisateurs de R venant chercher du savoir-faire logiciel sur la mise en œuvre des analyses statistiques « avancées » telles que l'ANOVA mono et multi-facteurs, la régression linéaire, les analyses multivariées (ACP, AFC...) et la classification (CAH). 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...)

PRÉ-REQUIS

- Il est **INDISPENSABLE** que l'apprenant ait les connaissances de base sur le logiciel R : création et manipulation des objets sous R, importation des données, utilisation simple de R en mode script ou en mode console, mise en œuvre de tests d'hypothèses tels que Student...
- Il est également **INDISPENSABLE** que l'apprenant ait des connaissances sur les différentes analyses statistiques mises en œuvre au cours de cette formation

THEMES PRINCIPAUX

- Création de script pour l'analyse des données
- Mise en œuvre d'une ANOVA mono-facteur et tests post-Hoc (comparaisons multiples) sous R
- Mise en œuvre d'une ANOVA multi-facteurs sous R (avec ou sans interactions, à mesures répétées, à facteurs imbriqués)
- Mise en œuvre d'une régression linéaire multiple sous R
- Mise en œuvre des méthodes d'analyse multidimensionnelle des données : ACP, AFC, CAH...

OBJECTIFS PEDAGOGIQUES

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

- Lire et écrire des scripts « avancés » d'analyse des données
- Mettre en œuvre et extraire les résultats aussi bien d'une analyse de la variance à 1 facteur que des tests post-Hoc de comparaisons multiples
- Mettre en œuvre et extraire les résultats d'une analyse de la variance multi-facteurs
- Mettre en œuvre et extraire les résultats d'une régression linéaire simple et multiple
- Mettre en œuvre et extraire les résultats d'une analyse multidimensionnelle des données : ACP, AFC, CAH...

QUELQUES MOTS-CLES

- Scripts
- ANOVA sous R
- Régression linéaire sous R
- ACP, AFC, CAH... sous R

OBJECTIFS OPERATIONNELS ET 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
- Affichage, sauvegarde, suppression et chargement des objets en mémoire
- Importation des données contenues dans un classeur Excel
- Installation et utilisation d'un package
- Utilisation d'interfaces type Rcmdr
- 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 d'observation ou de variables à un tableau de données
- Empilement / dépilement d'un tableau de données
- Conversion des objets et contrôle du type des objets
 - ✓ Commandes `is.something`
 - ✓ Commandes `as.something`
- Gestion des données manquantes
- Présentation et utilisation des fonctions de type `apply` : `apply`, `tapply`...
- Création de tableaux de synthèse (tableaux croisés)

Créer et exécuter des scripts avancés

- Exécution des scripts
 - ✓ dans la console R
 - ✓ à partir d'un autre script
- Création de scripts personnalisés
- Utilisation des scripts pour l'automatisation des analyses

Revoir la mise en œuvre des analyses statistiques de base 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)
 - ✓ Comparaison des variances (Fisher)
 - ✓ Comparaison de proportions (χ^2)
 - ✓ 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 et estimation des coefficients du modèle
 - ✓ Vérification des conditions d'utilisation du modèle : Normalité et homogénéité des dispersions (test de Bartlett, test de Levene)
 - ✓ Représentation graphique des résultats de l'analyse
 - ✓ Tests post-Hoc (comparaisons multiples)
 - Les commandes de base (TukeyHSD, pairwise.t.test)
 - Package spécifique (multcomp)
 - ✓ Tests non paramétriques : Kruskal-Wallis, Friedman
- Mise en œuvre d'une ANOVA multi-facteurs
 - ✓ Éléments de syntaxe : *, |, /, Error...
 - ✓ Effets simples, effets croisés
 - ✓ ANOVA à mesures répétées
 - ✓ ANOVA à facteur imbriqué
 - ✓ Graphique des interactions
- Somme des carrés de type I / somme des carrés de type III

Mettre en œuvre une régression linéaire sous R

- Commande summary() / commande 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
 - ✓ Normalité des résidus
 - ✓ Homogénéité des résidus
- Prédiction et intervalles de confiance d'une prédiction
- Diagnostic des problèmes de multicollinéarité
- Régression linéaire pas à pas

Réaliser une analyse factorielle multivariée et une classification sous R

- Présentation des principaux packages
- ACP (Analyse en Composantes Principales)
 - ✓ Réaliser une ACP sous R
 - ✓ Valeurs propres et diagramme des valeurs propres
 - ✓ Contributions et cosinus² des variables
 - ✓ Contributions et cosinus² des individus
 - ✓ Cercle factoriel
 - ✓ Graphique des individus
- AFC (Analyse Factorielle des Correspondances)
 - ✓ Profils-lignes et profils-colonnes
 - ✓ Réaliser une AFC sous R
 - ✓ Valeurs propres et diagramme des valeurs propres
 - ✓ Coordonnées, cosinus² et contributions des lignes, des colonnes
 - ✓ Représentation des modalités dans le plan factoriel
- CAH (Classification Ascendante Hiérarchique)
 - ✓ Commandes hclust(), kmeans() et hcpc()
 - ✓ Création de la matrice des distances
 - ✓ Dendrogramme
 - ✓ Récupération des classes

Mettre en œuvre d'autres analyses statistiques (sur demande et si le temps le permet)

La mise en œuvre d'autres analyses statistiques « avancées » des données pourra être réalisée sur demande au cours du stage si le temps le permet. Par exemple, nous pourrions aborder la mise en place d'analyses statistiques telles que :

- L'AFCM, l'ACC, l'AFD... (analyses factorielles multivariées et méthodes de classements)
- La régression PLS
- La régression logistique
- Les modèles linéaires généraux (ANCOVA...)
- ...

Méthodes et moyens :

- Explications théoriques suivies de pratiques guidées puis mises en autonomie
- 1 vidéoprojecteur par salle
- 1 ordinateur par stagiaire

Méthodes d'évaluation des acquis :

- Exercices d'applications
- Evaluation de fin de stage

Profil formateur :

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de compétences aussi bien au niveau technique que pédagogique.

Support stagiaire :

- Support électronique (dématérialisé)
- Les exercices d'accompagnement peuvent être récupérés sur clef USB