



PROJET RISE

Recherche, Innovation, Surveillance et Évaluation

TERMES DE REFERENCE

Formation : Initiation à RStudio pour l'analyse des données de santé

Durée de la formation : 6 sessions (3h le matin et 3 h l'après-midi)
Dates de la formation : 1-3-5-9-10-12 mars 2021
Lieu de formation : Institut Pasteur de Madagascar

1 Cadre de la formation

Les données générées au sein des systèmes de santé, données de surveillance, de laboratoire, données hospitalières sont en général sous-exploitées. Or, leur exploitation est un levier d'amélioration pour la prise de décision, pour identifier les programmes prioritaires et améliorer la santé des populations. La valeur des données s'impose au-delà de leur volume et provient surtout de la manière dont elles sont exploitées et, bien entendu, des enjeux en présence.

RStudio est un environnement de développement intégré libre, gratuit, et qui fonctionne sous Windows, Mac OS X et Linux. Plus particulièrement, il comporte des moyens qui rendent possibles la manipulation des données, les calculs et les représentations graphiques. RStudio complète R, un logiciel dédié aux études statistiques. Il offre un langage de programmation complet et bénéficie du support d'une large communauté de développeurs. RStudio possède:

- un grand nombre d'outils pour l'analyse des données et les méthodes statistiques,
- des moyens graphiques pour visualiser les analyses, et
- un langage de programmation simple et performant comportant : conditions, boucles, moyens d'entrées sorties, possibilité de définir des fonctions récursives.

Il est parfaitement adapté à l'analyse des données de santé.

2 Objectif

L'objectif principal de la formation est de renforcer la connaissance et les compétences des participants à l'utilisation de RStudio pour l'analyse des données de santé.

L'analyse des données de santé publique consiste en la collecte, l'analyse et la dissémination des résultats afin d'apporter de l'information qui puisse être utilisée pour monitorer et améliorer la santé des populations.

RStudio est une suite intégrée d'outils facilitant la manipulation de données, les calculs statistiques et la représentation graphique des résultats. Il inclut :

- Des outils permettant de facilement manipuler et sauvegarder les données
- Une suite d'opérateurs pour réaliser des calculs de toute sorte

-
- Une collection très large d'outils pour l'analyse intermédiaire et avancée, y compris un large éventail de tests statistiques
 - Des outils graphiques et géographiques pour l'analyse des données pour affichage à l'écran et sauvegarde sur tous types de documents (Word, PDF, Excel...)
 - Une programmation simple et efficace incluant les sauts conditionnels, des choix larges en matière d'import de données.

3 Participants ciblés

La formation est prévue pour un maximum de 12 participants.

Les prérequis sont :

- Une expérience avérée dans l'analyse de données de santé (Excel ou autre logiciel)
- Un bon niveau de pratique et de connaissance en informatique
- Une implication dans l'analyse de données de santé dans une institution publique, un laboratoire, un institut de recherche ou toute organisation en lien avec la santé
- Une expérience avérée en épidémiologie ou en santé publique.

4 Approche méthodologique

La formation sera organisée autour d'études de cas et des exercices pratiques à réaliser à partir de base de données qui seront fournies aux participants.

Il s'agit à ce stade d'une initiation au logiciel RStudio.

5 Organisation de la formation

Il y aura environ 30 % de présentations théoriques et 70% d'exercices pratiques sur le logiciel RStudio.

Les cours seront dispensés par 2 intervenants d'Epiconcept :

- Dr Loïc Flachet, médecin épidémiologiste, directeur opérationnel d'Epiconcept Océan Indien basé à Maurice
- Mme Lore Merdrignac, Epidémiologiste, bio informaticienne

Au début et à la fin de la formation, une évaluation du niveau de connaissance des participants sera réalisée.

La formation sera organisée comme suit :

Journée 1

Présentation des interfaces et modules de RStudio :

- Présentation des interfaces
- Travailler et s'organiser avec RStudio : répertoires de travail, création d'un script d'analyse, création de listing de données, suppression des objets dans R
- Présentation des packages R pour l'analyse descriptive de données et des principales commandes et fonctions
- Travailler avec une base de données, lecture, import, exploration, manipulation, exportation d'une base de données

Journées 2 à 5

Les participants seront amenés à réaliser une analyse complète de la qualité des données et une analyse descriptive à partir d'étude de cas préparés spécialement pour la formation. Ils utiliseront les capacités de programmation de RStudio pour cela:

- Analyse de la qualité des données : analyse des données manquantes et des données aberrantes, nettoyage de la base de données
- Analyse descriptive: présentation des résultats sous forme de tableaux et de graphiques analysant les données en termes de temps, lieu, personne en utilisant les packages R les plus appropriés

Journée 6

Durant la dernière journée de la formation les participants apprendront à générer des rapports automatisés en utilisant les fonctions R (rapports sous format PDF, html, Word).

6 Résultats attendus

A la fin de la formation les participants seront capables de :

- faire une analyse de la qualité des données avec RStudio
- réaliser une analyse descriptive des données
- utiliser les principales fonctions de présentation de données de RStudio
- générer des rapports automatisés sous différents formats