



Ecole d'été de Santé Publique et d'Epidémiologie de Paris-Saclay



R. Bauer et R. Foare

4. Initiation à l'analyse de données avec R (introduction au tidyverse)

Chaque après-midi du 22 juin au 26 juin 2026

Objectifs

Etre autonome avec le logiciel R pour l'importation d'un tableau de données, la génération de variables, la réalisation de statistiques descriptives, et des tests et modélisations les plus courants en épidémiologie et recherche clinique.

Le but de ce module n'est pas de discuter des résultats d'analyses statistiques mais d'apprendre à utiliser les commandes pour réaliser ces analyses.

Public concerné

Le cours est prévu pour les « vrais débutants » en programmation R. En revanche, des bases théoriques sont requises : théorie des tests, principaux tests univariés, régression linéaire

Pré-requis

- Connaissances théoriques en Biostatistique et Epidémiologie
 - au minimum Licence pro en statistique ou M1 de Santé Publique (CESAM ou équivalent)
 - idéalement M2 de Santé Publique - Epidémiologie ou Recherche clinique
- Connaissances pratiques :
 - Environnement Windows : pratique courante indispensable.

Programme

- Prise en main du logiciel R et de l'environnement RStudio
- Manipulation d'une base de données : Importation d'une table, tri sur une variable, sélection de lignes selon des conditions définies, fusion de plusieurs tables, génération de nouvelles variables, etc.
- Production de statistiques descriptives simples et de graphiques

La manipulation de base de données et la production de graphiques seront présentées avec l'extension *tidyverse*. L'extension *tidyverse* permet d'effectuer un grand nombre d'opération courantes et vise à faciliter le travail de nettoyage et de préparation de bases de données, préalable à l'analyse statistiques.

Modalités pédagogiques

Cours en salle d'informatique.

Un stagiaire par ordinateur.

Possibilité d'utiliser son ordinateur personnel, avec le logiciel R et l'environnement RStudio préalablement installés (installation gratuite).