

# Traitement Statistique avec R

## INTRODUCTION

Ecole Nationale de la Statistique et de l'Analyse Economique  
ENSAE Pierre Ndiaye  
ISEP2

M. H. DIALLO  
[hadyafia@gmail.com](mailto:hadyafia@gmail.com)  
(+221) 778441560

13 mars 2023



# Introduction



# Avantages

Voici quelques-uns des avantages :

- R est gratuit et constitue le premier avantage majeur.
- Vous pouvez installer R sur autant d'ordinateurs que vous le souhaitez
- R est facile à collaborer et à apprendre
- Tout le monde peut écrire un package complémentaire pour R
- Fournit l'analyse la plus avancée
- R a la capacité de recherche reproductible. Un projet peut être répété facilement lorsque de nouvelles données sont disponibles.

# Qui est le bon public pour apprendre le langage de programmation R ?

## ① Data Analyst

R serait l'outil idéal lorsque vous serez prêt à développer un métier dans l'Analyse de données et souhaitez vous imprégner du langage de codage open source fabriqué par les Statisticiens à destination des Statistiques.

## ② Data Scientist

- Codage R pour la science des données
- Chargement de données à partir d'un fichier ou simplement d'une base de données.
- Des requêtes courantes sur les données comme un résumé, des diagrammes de dispersion, des diagrammes de boîte, etc.
- Traitement des données, apurement, réparer les données manquantes.
- Séparer les données
- Faire un modèle centré et des prévisions.
- Valider les résultats.
- Visualisation des données

# Plan du cours

- 1 Introduction à R : les bases
  - Pré-requis : Installation
  - Type de données
  - Les opérateurs dans R
- 2 Manipulation de données : les bases
  - Analyse statistique de base
  - Visualisation de données
- 3 Programmation en R
- 4 R pour les Data Scientists
  - Manipulation des données avec Tidyverse
  - Manipulation et analyse des données
  - Définition des packages
- 5 Communication des résultats
  - Rédaction d'un rapport automatisé
  - Création d'un dashboard
- 6 Exposés
- 7 Contrôle de connaissances

# Type de données

- 1 Jeu de données
- 2 Objets
- 3 Vecteurs
- 4 Facteurs
- 5 Matrices
- 6 Liste
- 7 Data Frame

# Les opérateurs dans R

- ① O. Arithmétique
- ② O. relationnel
- ③ O. Logique
- ④ O. d'assignation

# Manipulation de données : les bases

- 1 Importation
- 2 Structure des données
- 3 Renommer, créer, labeliser les variables
- 4 Récoder ;
- 5 changer de type ;
- 6 découpage en classe ;
- 7 Répérer les Valeurs manquantes ;
- 8 Détecter les individus aberrants ;
- 9 Exportation ;
- 10 Transformer les bases ;



# Programmation en R

- 1 Fonctions
- 2 Structure if
- 3 Structure for
- 4 Structure while

# Manipulation des données avec Tidyverse

- ① L'utilisation des pipe
- ② Visualisation
- ③ Manipulation des tableaux de données
- ④ import/export de données
- ⑤ manipulation de variables
- ⑥ extraction de données du Web
- ⑦ programmation

# Manipulation des données (suite)

- 1 Analyse graphique
- 2 Statistique bivariée
- 3 Inférence statistique avec R
- 4 Test paramétrique et non paramétrique
- 5 Regression
- 6 Anova
- 7 Econométrie des Var qual
- 8 Analyse des données multidimensionnelles

# Définition des packages

- 1 Utilisation des packages
- 2 Création et mise en ligne des packages

# Rédaction d'un rapport automatisé

- 1 Élément d'un document R markdown /R sweave (intro to html)
- 2 Personnalisation d'un document R markdown
- 3 Présentation des tableaux

# Création d'un dashboard

- Introduction sur R shiny
- Creation de dashbord
- creation d'une application web
- Interaction de R avec les outils d'analyse de données (STATA, Python, SAS...)

1. Les méthodes d'imputations des données
2. Visualisation des données avec la bibliothèque ggplot2
3. Traitement des données
4. Utilisation de la bibliothèque Shiny pour la création d'applications web interactives en R.
5. Analyse de survie avec R
6. Data mining avec R
7. Sondage avec R
8. Modélisation de séries chronologiques avec R.
9. Traitement des données spatiales avec R
10. Utilisation de R pour la reconnaissance de formes et l'apprentissage automatique.

# Contrôle de connaissances

- 1 Contrôle écrit
- 2 Projet



# Références bibliographiques

- 1 Programmation et analyse statistique avec R, Ellipse Edition, Christian Paroissin, 2015
- 2 Pierre-André Cornillon et al. (2012). Statistiques avec R, 3e édition revue et augmentée, Presses universitaires de Rennes.
- 3 Pierre Lafaye de Micheaux, Rémy Drouilhet, Benoît Liquet (2014). Le logiciel R :
- 4 Maîtriser le langage – Effectuer des analyses statistiques, 2e édition, Springer.
- 5 Crawley, M.J. (2013). The R book, 2e edition, Wiley.
- 6 Introduction à la programmation en R, 5ème édition, Vincent Goulet et Laurent Caron, 2016
- 7 Introduction to data science with R

# Méthode de travail

- Learning by doing !
- Exercices d'applications
- Homework
- dossier de travail : [Trouver ici les fichiers du cours](#)
- [Remplissez ce formulaire](#)

# Let's go !

