



FORMATION

ouverte aux doctorant·es de l'École doctorale
Sciences Sociales et Humanités

STATISTIQUES AVEC LE LOGICIEL R POUR DÉBUTANT·ES

PROGRAMME ET INSCRIPTION



Inscription
obligatoire,
dans la limite des
places disponibles

**10 &
11 juin
2025**

Durée de la formation : 9 heures

De 10h à 17h le 10 juin,
de 10h à 13h le 11 juin.

► **Formation assurée par Mathieu Ferry,**
maitre de conférences, laboratoire Printemps (UVSQ/CNRS)



UVSQ
Bâtiment Vauban - Salle 434, 4^{ème} étage
47, boulevard Vauban - 78280 Guyancourt

CONTACT

Émeline Caulet
emeline.caulet@universite-paris-saclay.fr

**université
PARIS-SACLAY**

GRADUATE SCHOOL
Sociologie et
Science politique

Statistiques avec R pour débutant·es Descriptif indicatif de la formation

Contenu et objectifs

La mobilisation de données quantitatives trouve un regain particulier ces dernières années dans le cadre d'enquêtes par « méthodes mixtes » et grâce à l'essor de nouvelles données (« open data », « big data »), en plus des sources statistiques plus classiques (données administratives, enquêtes par questionnaire). Cependant, les savoir-faire requis rebutent parfois les chercheur·euses. L'apprentissage du langage R et de ses nombreuses « librairies » peut décourager même les utilisateur·rices les plus motivées, en particulier face aux inévitables « errors » et « warnings » affichés par la console RStudio. Cette formation vise à introduire au langage R, aux outils indispensables d'analyse de données pour les sciences sociales et à quelques « bonnes pratiques » de gestion et d'exploitation des données.

La formation est organisée en trois séances de trois heures, qui se déroulent sous un même format. Les différents points de connaissance sont illustrés par de courts scripts à compléter, corriger ou rédiger. Des travaux pratiques seront réalisés lors de chaque demi-journée de formation, à partir de données issues de sources variées.

En particulier, la formation propose de présenter les librairies et les fonctions récurrentes de gestion des données avec les outils du « tidyverse », de présenter les outils pour réaliser des statistiques univariées et bivariées et leur mise en forme sous forme de tableaux ou de graphiques. La dernière session s'intéresse à un type de données particulier : les agrégats de textes (corpus) dont on peut analyser des « volumes » importants grâce à l'analyse textuelle exploratoire.

À l'issue de la formation, chaque participant·e sera en mesure d'exploiter des sources statistiques pertinentes pour son objet de recherche et d'en réaliser des analyses statistiques simples sous R.

Programme des séances

Séance 1 – 10 juin 2025, 10h à 13h.

- 10h-11h. Introduction. Le langage R et ses multiples dialectes. Prise en main et premiers éléments de langage
- 11h-11h30. Travaux pratiques
- 11h30-12h. Statistiques univariées (indicateurs de centralité et de dispersion)
- 12h-13h. Travaux pratiques

Séance 2 – 10 juin 2025, 14h à 17h.

- 14h-14h30. Statistiques bivariées : tableaux croisés et moyenne par groupe
- 14h30-15h30. Travaux pratiques
- 15h30-16h. Statistiques multidimensionnelles et régressions
- 16h-17h. Travaux pratiques

Séance 3 – 11 juin 2025, 10h à 13h.

- 10h-11h. Introduction à l'analyse statistique textuelle et construction d'un corpus de presse exploitable grâce à Europresse et R.
- 11h-12h. Travaux pratiques.
- 11h30-12h. Présentation d'outils d'exploration d'un corpus textuel.
- 12h-13h. Travaux pratiques.